Sustainable Cities and Society 72 (2021) 103054

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

A novel performance indicator for the assessment of the learning ability of smart buildings

Kari Alanne

Department of Mechanical Engineering, School of Engineering, Aalto University, PO Box 14400, FI-00076, Aalto, Finland

ARTICLE INFO

Keywords:
Smart building
Building intelligence
Learning
Performance assessment

ABSTRACT

The rapid development of artificial intelligence (AI) and machine learning (ML) has made it topical to consider learning ability as one of the key performance characteristics of buildings. So far, the buildings' learning ability has not explained or clarified by definitions or in terms of the proposed frameworks of key performance indicators (KPI). In this paper, a novel performance indicator based on the concept of learning gain is developed to quantify the learning ability of buildings by way of a single, dimensionless number between zero and unity. The implementation of the new Learning Ability Index (*LAI*) is demonstrated by way of three different case studies chosen from the literature. It is concluded that *LAI* is an easy and illustrative tool to assess the learning ability of buildings. Particularly, it is useful for monitoring the performance of data-driven processes, when pursuing the preferred strategies to reach higher levels of building intelligence. The *LAI* considers the time invested in learning plus the quality and diversity of learning material. It is flexible with respect to system boundaries or the performance metrics, wherefore it can be implemented as a generic indicator of system evolution, as well.

1. Introduction

The need to associate the word 'learning' with buildings originates from two interlinked undercurrents of sustainable development, namely, the intention to help the communities' energy system cope with an increased penetration of intermittent renewable energy resources (solar, wind) in tandem with tighter demand of energy efficiency and resilience, and to equip buildings with an ability to dynamically respond to changing boundary conditions (e.g. user's needs, changing climate and fluctuating grid prices) (EPBD Recast (2010), Al Dakheel, Del Pero, Aste, & Leonforte et al. (2020)).

The qualities of learning (Karpook, 2017) and cognition (Xu, Lu, Xue, & Chen, 2019) have been often mentioned among the buildings' 'intelligent' and 'smart' qualities. Here, 'building intelligence' refers to the diffusion of ICT in the infrastructure and the increasing the ability of artificial systems to operate autonomously, whereas 'smartness' entails the building's ability to interact with people and community (Wang et al. (2020); Albino, Berardi, and Dangelico (2015)). Again, the buildings' ability to learn is referred to as their capability to adapt to occupants' preferences and behaviour, occupancy patterns, productivity, indoor environmental preferences, and thermal behaviour of the building and its environment (Mofidi and Akbari (2020); Lê, Nguyen, and Barnett (2012)).

In principle, buildings' learning is always data-driven, and it is based on the analysis of simulated or measured data (e.g. Jazizadeh, Ghahramani, Becerik-Gerber, Kichkaylo, and Orosz (2014); Marinakis, Karakosta, Doukas, Androulaki, and Psarras (2013)). Machine learning algorithms integrated into the building energy management system (BEMS) allow a building to i) observe its own status (e.g. Araya, Grolinger, El Yamany, Capretz, & Bitsuamlak, 2017), ii) predict changes (e.g. Cao et al., 2020), iii) adjust the operation (e.g. Azuatalam, Lee, de Nijs, & Liebman, 2020), iv) solve problems (diagnose and correct faults) (e.g. Li, Zhou, Hu, & Spanos, 2016), and ultimately v) interact with occupants (e.g. Konstantakopoulos et al., 2019) in reactive, real-time, predictive or proactive manner.

In brief, the learning ability can be defined as a building's ability to improve its intelligent, smart, or cognitive performance (proficiency) over time on the basis of accumulated experience (e.g. training data).

Abbreviations: AI, Artificial Intelligence; ALR, Average Learning Rate; ASHRAE, American Society of Heating, Refrigerating and Air-Conditioning Engineers; BEMS, Building Energy Management System; BI, Building Intelligence; BIQ, Building Intelligence Quotient; BLE, Bluetooth Low Energy; DT, Digital Twin; EPBD, Energy Performance of Buildings Directive; HVAC, Heating, Ventilation and Air-Conditioning; IBMS, Intelligent Building Management System; ICT, Information and Communication Technologies; IPMVP, International Performance Measurement & Verification Protocol; KPI, Key Performance Indicator; LAI, Learning Ability Index; LM, Linear (Regression) Model; LR, Learning Rate; MBE, Mean Bias Error; ML, Machine Learning; NLG, Normalized Learning Gain; NSGA, Non-Dominated Sorting Genetic Algorithm; RF, Random Forest; RMSE, Root Mean Square Error; SS, Sample Size.

https://doi.org/10.1016/j.scs.2021.103054
Received 8 December 2020; Received in revised form 14 April 2021; Accepted 25 May 2021
Available online 28 May 2021

Good learning ability is an ability to implement the potential that enable learning efficiently and with high quality. To ensure the flawless and sustainable system operation in the conditions of rapidly increasing level of automation, the buildings' learning ability should be monitorable, measurable and assessable. Again, the assessment protocol should be straightforward enough and compatible with generic performance assessment frameworks.

An attempt to quantify the learning ability of buildings is reported by Volkov (2013) in terms of Building Intelligence Quotient (BIQ). The concept is further explained by Batov (2015), who point out that the BIQ definition should include the accuracy, learning speed and latency metrics of processes participating in observing the building parameters and their control. Roughly, the processes are classified into AI-initiated and human-initiated, which makes visible the contribution of AI in the management of the building and its systems. Another similar indicator is 'Consumer Engagement', which can be used to measure the involvement of users in the control over the energy use in the building (SCIS, 2017). The challenge with the use of aforementioned approaches is that due to the probabilistic nature of machine learning algorithms and complexity of a building as a system, it is difficult to obtain appropriate input data of various processes for the assessment. Hence, the assessment rather focuses on separate processes than the building as a holistic system.

The scientific contribution of the present study is to develop a novel indicator particularly for monitoring purposes and strategic optimization to measure buildings' (and in generic terms any complex system with AI-initiated processes) learning ability. Here, the proposed indicator for learning ability is the Learning Ability Index (*LAI*), which is a single, dimensionless number between zero and unity. The *LAI* encompasses the calculation of normalized learning gain (*NLG*) for two points in time plus the correction of the result by a coefficient to address the quality and diversity of the learning material. The use of *LAI* is explained through three (3) case studies covering various types of AI-initiated processes, performance indicators, data sources and temporal dimensions. Section 2 includes the theoretical premises of the calculation method. The case studies are presented in Section 3 and the findings of the work are discussed in Section 4.

2. Description of the Learning Ability Index (LAI)

2.1. Theoretical background, concepts and definitions

The first assumption for the assessment of a building's learning ability is that the (building) intelligence (BI) can be defined as a function of key performance indicators (KPIs), which, in turn, are functions of time as shown in Eq. (1):

$$BI(t) = f(KPI_1(t)\ldots KPI_n(t)) \tag{1}$$

Learning is quantified as the growth of the building intelligence function (Eq. (1)) over time. The lesser experience and the shorter time are required, the better is the building's ability to learn. Hence, the momentary learning rate (aka learning speed) can be defined as:

$$LR = \frac{dBI}{dt} \tag{2}$$

The learning rate (LR) is essentially an indicator of the effectiveness of learning, but it is not experienced useful in the given context as such, however, because it has a case-specific unit and it may obtain any (case-specific) values, which makes the interpretation of analysis results difficult. On the other hand, the learning rate may vary significantly during the learning process.

Therefore, the proposed Learning Ability Index (*LAI*) is developed on the basis of the concept of learning gain, which is defined as 'the difference between the skills, competencies, content knowledge and personal development demonstrated at two points in time' (McGrath, Guerin, Harte, Frearson, & Manville, 2015). Essentially, the learner (smart building) engages with a learning process and the learning ability interacts with case-specific KPIs to produce a learning-gain function which determines how prior-proficiency will map to post-proficiency (Piech, Bumbacher, & Davis, 2020).

Since there is no fixed definition for building intelligence, the number and type of KPIs in the intelligence function (Eq. (1)) is always case-specific. In general, the assessment may still rely on some established KPI frameworks. Al Dakheel et al. (2020), for example, present a table of 34 KPIs, where the KPIs have been divided into four sub-sets (nearly zero-energy targets, flexibility, monitoring, interaction with users). All in all, it is reasonable to seek such KPIs that enable an explicit and affordable metering of a building's smart performance instead of measuring the presence or involvement of services or technologies, which seems to be a prevailing approach in many existing assessment frameworks (e.g. Vigna, Pernetti, Pernigotto, and Gasparella (2020); Verbeke et al. (2017); Volkov (2013).

A practical approach for assessing the learning ability of smart buildings has been presented, for example, by Candanedo, Feldheim, and Deramaix (2018) who implemented linear regression (LM) and random forest (RF) algorithms for predicting average indoor temperatures on the basis of incomplete data. In their study, the Root Mean Square Error (RMSE) was chosen as the KPI, which quantifies the proficiency of the prediction model during the entire training period. Their study will be discussed with more details in Section 3.1.

2.2. Mathematical representation of the LAI

The starting point for the calculation of *LAI* is determining the value of the target building's intelligence function at two points in time so that the length of time interval between these moments is a pre-defined reference time period (t_{ref}) during which the learning process has been expected to take place.

Let BI_{pre} represent the building's intelligence in pre-learning phase (at the beginning of the reference period) and BI_{post} that in post-learning (monitoring) phase (at the end of the reference period). With the above definitions, the Average Learning Rate (*ALR*) within the reference period would be

$$ALR_{t_{ref}} = \frac{BI_{post} - BI_{pre}}{t_{ref}} \tag{3}$$

To obtain a single, dimensionless number within the range between zero and unity, the normalized learning gain (*NLG*) is first calculated so that baseline (representing $NLG = 0$) is fixed to the value of the building intelligence function (Eq. (1)) at the beginning of the reference period (BI_{pre}). Again, a pre-defined learning target (BI_{target}) is established with an assumption that if this target is obtained within the given reference time period, $NLG = 1$. Hence, the *NLG* over the reference time period is defined as:

$$NLG_{t_{ref}} = \frac{BI_{post} - BI_{pre}}{BI_{target} - BI_{pre}} \tag{4}$$

In other words, the *NLG* indicates the percentage/ratio of the obtained intelligence (by the reference time) and the theoretical intelligence or pre-defined learning target in comparison with a pre-defined baseline.

As the intelligence function (Eq. (1)) implies, the learning process may enable several individual learning targets (e.g. Han, May, Zhang, & Jin, 2020), which are assigned to each KPI_i and monitored at the beginning and the end of the reference time period. Thus, the *NLG* over the reference period is:

$$NLG_{t_{ref}} = f(NLG_1, \ldots, NLG_i, \ldots, NLG_n)_{t_{ref}} \tag{5}$$

where the *NLG* for the *i*-th KPI is

$$NLG_i = \frac{KPI_i - KPI_{i,0}}{KPI_{i,\infty} - KPI_{i,0}} \tag{6}$$

In Eq. (6), the subscript “0” refers to the baseline (pre-learning) and the subscript “∞” to the pre-defined learning target.

Despite the presence of several KPIs (learning targets), the *NLG* yet should remain as a dimensionless number between 0…1. Because all the single *NLG*s have been defined as proportional to their individual baselines and target values, they can be treated commensurate. Again, the holistic essence of the building intelligence allows that poor learning with respect to some KPI can be compensated (at least to some extent) by good learning with respect to another KPI. Therefore, the indicator-specific *NLG*s can be aggregated by a simple additive rule. The suggested aggregation rule is weighted average that produces the building’s *NLG* over all the KPIs as a single number as follows:

$$NLG_{t_{ref}} = \sum_{i=1}^{n} w_i NLG_i \tag{7}$$

where w_i is a KPI-specific weight and $\sum_{i=1}^{n} w_i = 1$.

The *NLG* calculated from Eq. (7) depends on the quality and diversity of experience acquired to gain intelligence. Inversely, the building’s intelligence suffers from poor quality of learning material. Here, the expression ‘learning material’ may refer, for example, to the training data used to teach machine learning algorithms or survey data gathered from occupants. If the value $NLG = 1$ is obtained with respect to the given learning material, one has to be aware of its limitations. Therefore, it is recommended that the aggregated *NLG* (Eq. (7)) is corrected by an experience coefficient ($X \in [0,1]$) for a more realistic description of the buildings’ learning ability, which also considers the quality (Q) and diversity (D) of the learning material. To that end, the desired Learning Ability Index (*LAI*) is defined as

$$LAI_{t_{ref}} = X \cdot NLG_{t_{ref}} \tag{8}$$

where the experience coefficient $X = F(Q, D)$. Here, the quantity of the learning material as such is not decisive, but rather its accuracy and effectiveness of learning (learning gain over the reference time). (The data quality may be also assessed through the concept of data efficiency.)

In a conventional error analysis, an uncertainty (error) in one dataset is repeated as a systematic uncertainty (error) throughout the entire analysis. Assuming that the sources of uncertainty are independent from each other, the total uncertainty is the sum of uncertainties related to each independent variable. Mathematically, this relationship is known as total differential.

The expectation is that both the improved accuracy and increased diversity of training data enhances learning, and the data sources can be treated as independent with respect to each other. For each data source, a normalized quality indicator ($Q \in [0,1]$) is assigned to depict its quality (commonly: accuracy). The quality indicators are data-specific, wherefore the detailed calculation is demonstrated in Section 3 through case studies.

Again, a diversity weight ($D \in [0,1]$) is elicited to describe the significance of each data source among the entire domain (pool) of learning material. Since the major role of the diversity weight is indicate the extent of an individual data source among all the data sources relevant to building intelligence, the sum of diversity weights is not necessarily unity, but it is unity at maximum. This idea is illustrated in Fig. 1, where the surface of each bubble denotes the significance of the corresponding data source within the entire data domain (the surface of the big bubble).

It is suggested that the experience coefficient X would be calculated as the sum of the pre-defined normalized quality indicators of individual data sources weighted by the corresponding diversity weights. Again, this is a simple additive rule, where the qualitative superiority of one data source may compensate the qualitative inferiority of another data source. Furthermore, the loss of diversity also results in the degradation of data quality.

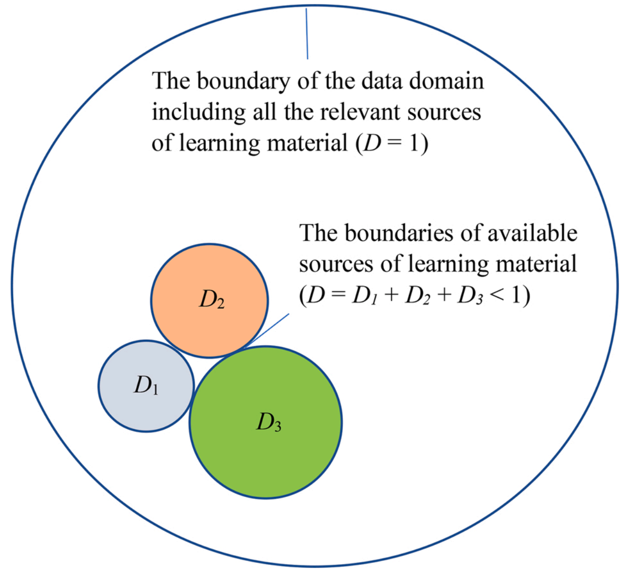

Fig. 1. Illustration of the set-up of diversity weights.

The experience coefficient is formulated as follows:

$$X = F(Q, D) = \sum_{j=1}^{m} D_j Q_j \tag{9}$$

where m is the number of data sources, D_j is the diversity weight and Q_j is the quality factor for the j-th data source.

The calculation procedure for the Learning Ability Index (*LAI*) in its entirety is depicted in Fig. 2. Here, the calculation starts from the bottom of the tree-structure (normalized *NLG*s by KPI, normalized quality indicators) and proceeds via the calculation of the aggregated *NLG* and the experience coefficient X to the *LAI*.

2.3. Analysis framework

The assessment of a building’s learning ability is a part of a cyclic process to pursue the most effective way to enhance building intelligence. Here, the starting point is choosing learning targets, i.e. determining the ‘skills’ the building is to achieve through learning (e.g. data management, prediction of changes, adjustment of the operation, problem solving). The learning process is an iterative loop in essence, where the implementation of learning method(s) and the assessment of their learning ability alternate with an expectation that the learning ability can be boosted by enhancing the quality of learning material and possibly changing the learning method. The learning ability of the previous assessment loop lays the baseline for the next one. The iteration loop may be endless, even though a case-specific stopping criterion may be set (e.g. the $LAI = 0.95$). Therefore, a brief discussion on the significance of the calculated *LAI*s is included in each loop. The assessment of building’s learning ability as an iterative loop is visualized in Fig. 3.

The calculation of *LAI* is involved in the process through the following three main steps:

Step 1: Inventory of learning methods, learning material and performance indicators

Step 2: Setting up the baseline(s), the utopia point(s) and reference time period(s)

Step 3: Calculation of *NLG*s, the experience coefficient, and *LAI*

The first step ‘inventory’ signifies not only the identification of appropriate data sources, but also possible data evaluation and management actions to support learning. For example, this may include measures such as the reconstruction of a set of training data or evaluation whether a survey data set is statistically significant or is it quality otherwise high enough. The second step entails stating the current status

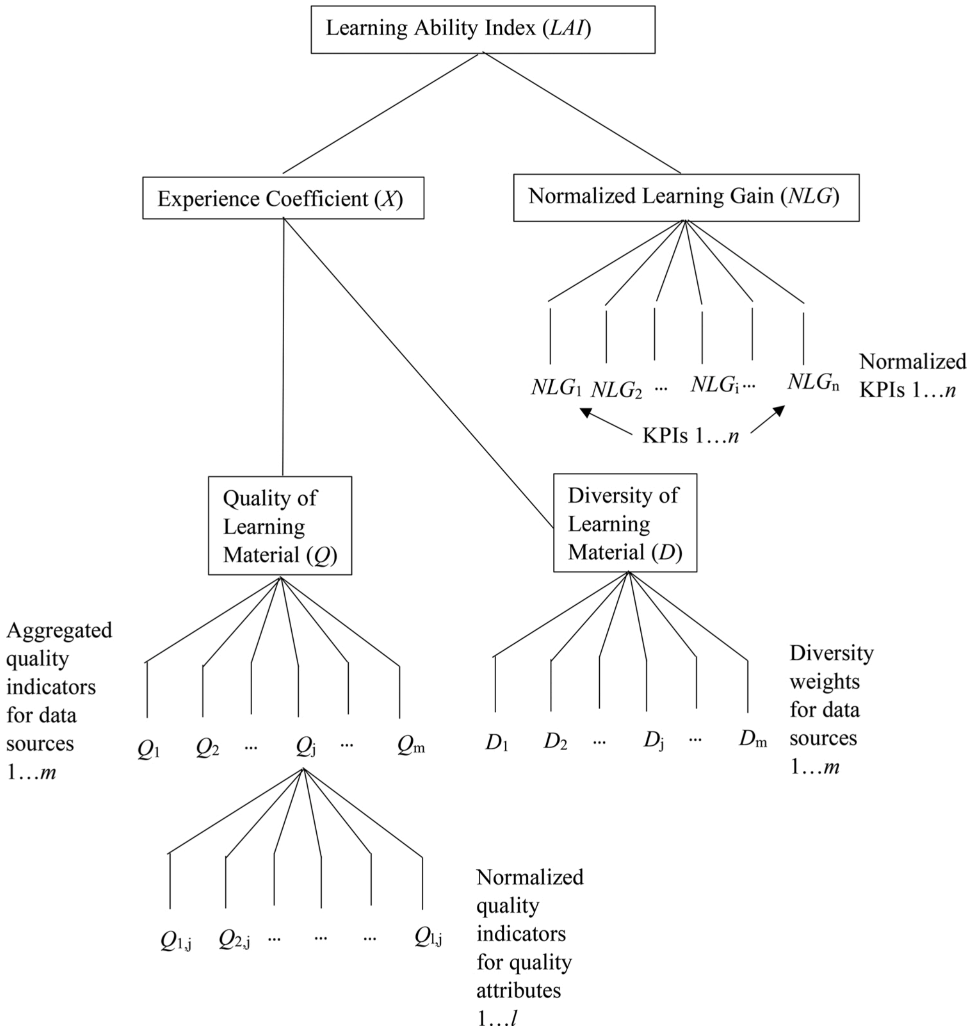

Fig. 2. The calculation of the Learning Ability Index (*LAI*) as a tree-shaped structure.

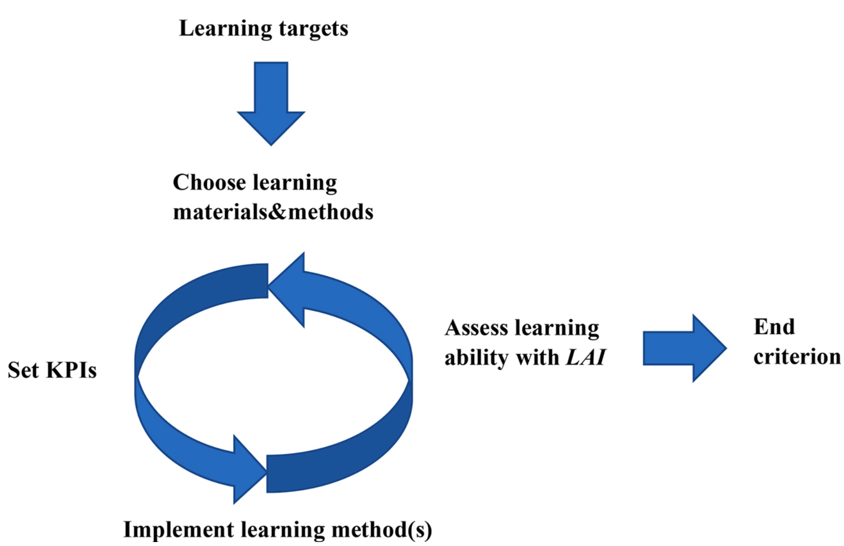

Fig. 3. Assessment of building's learning ability as an iterative loop.

and choosing the learning targets. Setting up the 'utopia points' (i.e. implicit learning targets) depends on the evaluator's ambitions, but yet it should be based some well-justified and realistic target level. The third step includes calculating the *NLG* from Eq. (7), the experience coefficient from Eq. (9), and the *LAI* from Eq. (8).

The assessment depicted in Fig. 3 is expected to yield the list of the recommended learning materials and methods to enhance the building intelligence in a particular case. Its results are case-specific, but since the output is a single number between zero and unity, the chosen learning strategy can be compared with corresponding ones in proportional terms. Again, the analysis framework and the calculation method can be generalized to any intelligent system where a reasonable domain of learning material and an appropriate framework of performance indicators exist. Hence, it may cover the assessment of learning abilities from the level of a unit process to that of the whole building stock. Here, it may be useful, for example, when outlining the overall transition strategies towards smart cities and communities.

The calculation of *LAI* is demonstrated in Section 3 through three case studies, which represent various types of learning material, performance indicators and learning methods.

3. Case studies

3.1. Case study 1: Data-driven reconstruction of incomplete indoor temperature dataset

Case study 1 has been selected with an aim to exemplify the calculation of the *LAI* on the basis of one performance indicator and quantitative learning material. The work of Candanedo et al. (2018) has been chosen as the reference, since it includes the option to a comparison between two learning methods, namely, the machine learning algorithms linear regression (LM) and random forest (RF). The performance indicator is root mean square error (RMSE), whereas the learning process is depicted as a reducing difference (residual) of the RMSE. The asset of this reference is an illustrative learning curve for both algorithms to graphically show the accumulated experience through learning in terms of reducing RMSE over time.

In the work of Candanedo et al. (2018), a calculation model for an average room temperature has been first calibrated (trained) using one dataset of measured temperatures (training data). Second, another dataset based on the same measurements has been used as a testing dataset to teach the model to predict missing temperature data and thus to complete the dataset. The training data has been collected during one (1) calendar year. The data set (including missing data) has 52704 entries for 2016 (at 10 min intervals), whereas the complete data set (free of missing data) has 37026 entries only, which corresponds to information for 257 days. The data acquisition phase as such could be described as a separate learning process. To that end, however, Candanedo et al. (2018) do not provide sufficient information. Hence, the *LAI* is calculated for the model training phase only.

The data provided by Candanedo et al. (2018) imply that the RMSE between the training dataset and the LM model is around 1.1 °C, whereas for the RF model it is around 0.4 °C. In other words, it appears that the RF method has performed better in the training phase, when it comes to finding the match between the training data and the model.

In this case study, learning is factually expressed by the models' increasing ability to predict missing temperatures in the testing phase. Here, both algorithms indicate their ability to match the test data by reducing the RMSE as close to the RMSE of training data as possible. In the very beginning of the testing phase (pre-learning), the RMSE between the test data and the trained LM model is around 5.7 °C, and it is 3.3 °C for the RF model. After the testing period (monitoring, post-test), the LM model stabilizes at the RMSE of around 1.17 °C having used the sample size (SS) of about 15300 for learning. Correspondingly, the RF requires the sample size of about 27300 to reach the RMSE (0.48 °C) in the testing set.

The best theoretical residual ('utopia point') is RMSE = 0 °C, even though none of the algorithms obtained it. This can be set as the learning target ($NLG = 1$) for both algorithms (LM&RF), anyway. Instead, the baseline is individual for each algorithm. If only the learning ability of the RF was assessed, the *NLG* would be zero-referenced to $\text{RMSE}_0 = 3.3$ °C which is the RMSE between the training data and the data predicted by the RF at the beginning (pre-test). Here, the baseline should be set to allow a comparison between the two algorithms within the range 0…1. Again, the better performance of the RF in the training phase (lower RMSE) should be rewarded rather than penalized. Therefore, the suggested way is to use the larger initial RMSE (5.7 °C) as the baseline for calculating the *NLG*.

The data provided by Candanedo et al. (2018) imply that the LM learns much quicker than the RF. Their data suggest, for example, that the LM only uses 2.51 s to learn from the sample size of 27300 data entries, whereas the RF needs 8549 s to that end. In other words, the LM algorithm is up to 8500 times faster to gain intelligence. Again, the LM needs only some 3.2 s to treat the data set of 34496 entries and to practically meet the best achievable RMSE (1.1 °C).

Candanedo et al. (2018) determine the elapsed time as a function of sample size (*SS*) using the following polynomial fits:

$$t_{LM} = 0.00006551 * SS + 0.740169 \quad (10)$$

$$t_{RF} = 0.00002665 * SS^2 + 0.06195 * SS - 169 \quad (11)$$

The performance data of the two algorithms during the first three seconds period of time calculated from Eqs. (10) and (11) is shown in Table 1. The data in Table 1 indicate, for example, that after the time period of 1 s, the LM algorithm has dealt with 3966 data samples and obtained the RMSE = 3.4 °C.

The reference time period (t_{ref}) in this type of an application is suggested to be chosen on the basis of application-specific data sampling intervals. Commonly, the range of sampling interval in whole-system analyses vary between 10 s…3600 s. In the present study, however, the time usage of the compared algorithms varies significantly. For practical reasons, the time period between pre-learning and monitoring is set to $t_{ref} = 3$ s in this study.

On the basis of the data in Table 1, the normalized learning gain ($t_{\text{ref}} = 3$ s) is adjusted into the scale 0…1 so that for the LM algorithm it is $NLG_{\text{LM,3s}}$= (5.7 °C – 1.1 °C) / (5.7 °C – 0.0 °C) = 0.81, whereas for the RF it is $NLG_{\text{RF,3s}}$ = (5.7 °C – 2.8 °C) / (5.7 °C – 0.0 °C) = 0.51. Correspondingly, the Average Learning Rate for the LM is $ALR_{\text{LM,3s}}$ = (5.7 – 1.1) °C/3 s = 1.5 °C/s and $ALR_{\text{RF,3s}}$ = (3.3 – 2.8) °C/3 s = 0.2 °C/s.

The normalized learning gains calculated on the basis of the RMSEs listed in Table 1 as a function elapsed time for both algorithms, are shown in Fig. 4. Since the starting level for the RF is higher ($NLG_{\text{RF,0s}} = 0.42$), the graph in Fig. 4 makes visible the fact that the RF algorithm has performed better in the first training phase (i.e. the RMSE of the training data is lower than that of the LM algorithm). Again, none of the methods is able to reach normalized learning gains higher than 0.8 in the frame of the given time reference.

To evaluate the experience coefficient *X*, the resolution and accuracy of the training data are chosen as the quality attributes since they have been quantitatively indicated by Candanedo et al. (2018). The realized sampling interval is 10 min = 1/6 h. Thus, the normalized quality indicator for data resolution is Q_R = (1 h – 1/6 h) / (1 h – 0 h) = 0.83. In other words, the normalization is zero-referenced to the calculation step of whole-system simulation (1 h), whereas the 'utopia point' is 0 h.

The reported accuracy of the temperature measurement is ±0.5 °C and ±3 % for the relative humidity (Candanedo et al., 2018). Assuming the average indoor temperature of 21 °C, the relative error would be ±2.4 % and, correspondingly, the proposed normalized quality indicator for the accuracy of the temperature measurement is its complement, i.e. $Q_A = 1 - 0.024 = 0.976$.

Table 1
The performance data for the LM and RF algorithms for the reference period.

Time [s]	SS_{LM} [-]	RMSE_{LM} [°C]	SS_{RF} [-]	RMSE_{RF} [°C]
0	0	5.7	0	3.3
1	3966	3.4	3943	2.8
2	19231	1.2	3949	2.8
3	34496	1.1	3956	2.8

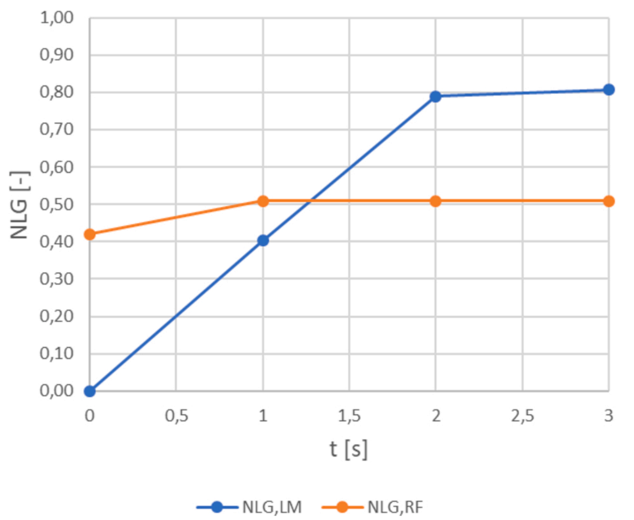

Fig. 4. The normalized learning gains for LM and RF algorithms as a function of time.

The relationship between the resolution and accuracy is considered additive, i.e. both a tight resolution and good accuracy bring added value to the quality of the training data. Assuming both quality attributes equal in weights ($w_R = w_A = 0.5$), the aggregated quality indicator can be calculated as the weighted average ($Q = w_R \cdot Q_R + w_A \cdot Q_A = 0.5 \cdot 0.83 + 0.5 \cdot 0.976 = 0.9$) and it is equal for both methods.

Again, the experience coefficient $X = X(Q,D)$ is formulated as $X = F(Q,D) = w_D D + w_Q Q$ and the Learning Ability Index is defined as $LAI = X \cdot NLG = (w_D D + w_Q Q) \cdot NLG$. Both algorithms (LM, RF) use temperature measurement as the only data source and there are no expectations concerning optional data sources in the domain, either, wherefore the diversity weight can set to unity. Thus, the experience coefficient $X = 0.5 \cdot 1.0 + 0.5 \cdot 0.90 = 0.95$ for both methods.

Finally, the values of LAIs are $LAI_{3s}^{LM} = 0.95 \cdot 0.81 = 0.77$ and $LAI_{3s}^{RF} = 0.95 \cdot 0.51 = 0.48$. The conclusion is that the LM method is preferred when the learning targets are described as above. The explanation is the LM method's extremely high learning rate, which repeals the benefits of the RF method in this application.

3.2. Case study 2: Data-driven optimization of energy consumption and occupant comfort

Case study 2 has been selected with an aim to exemplify the calculation of the *LAI* on the basis of two performance indicators and quantitative learning material. The work of Salimi & Hammad (2020) has been chosen as the reference, since it implements an integrated, demand-driven optimisation method to simultaneously improve the performance of an office building with respect to two KPIs, namely, indoor comfort and annual energy consumption. Here, learning signifies ending up with a conclusion on the most preferred control strategy (including set-points for heating, cooling and illuminance) for a BEMS system.

Salimi & Hammad (2020) introduce the two KPIs by way of a generic performance indicator (P), which may refer to either i) the annual building energy consumption or ii) the hours outside the ASHRAE 55 comfort regions (i.e. the number of discomfort hours). The change of the generic performance indicator due to optimization (ΔP) is calculated from Eq. (12):

$$\Delta P = \frac{P^c - P_0^{unc}}{P_0^{unc}} \tag{12}$$

where P^c is the value of energy consumption or the number of discomfort hours after the optimization (monitoring phase). The performance without the optimized control strategy denotes the baseline (pre-learning phase) and it is indicated as P_0^{unc}.

The target building is a realistic, open-space office (35.1 m^2). Real occupancy data has been collected over the period of one year (from April 1 st, 2017 to March 31 st, 2018). Bluetooth Low Energy (BLE)-based monitoring system has been used for data acquisition with the temporal resolution level of one (1) second. To determine the annual space energy consumption and discomfort hours, Salimi & Hammad (2020) use the EnergyPlus (v.8.6) whole-building simulation software with the simulation time step of 1 min. Here, set-points temperatures for heating, cooling and illuminance according to either a simple or detailed scheme are adjusted either every 30 min or every 60 min, i.e. the control resolution is either 30 min or 60 min. A separate optimization algorithm (Non-Dominated Sorting Genetic Algorithm, NSGA-II) feeds the control variables to the simulation tool (EnergyPlus) and runs the tool with annual simulation and a set of control variables until a pre-defined stopping criterion is fulfilled. In the work of Salimi & Hammad (2020), the NSGA-II algorithm was run for 100 generations with a population size of 20.

Salimi & Hammad (2020) introduce six (6) alternative control strategies (Cases) in total. Case 1 is based on standard occupancy, wherefore it is omitted from the present study. Instead, Case 2 is the non-optimized control based on the real occupancy schedule, and it is acquired as the baseline for the calculation of the *NLG*. Cases 3 and 5 are optimized control strategies with the HVAC control resolution of 30 min and Cases 4 and 6 the control resolution of 1 h. Here, they represent the system's status after the learning process. The reader is referred to the original research article for further details.

According to Salimi & Hammad (2020), the simulated baseline energy consumption (Case 2) is 6266 kW h (179 kW h/m^2), whereas the number of discomfort hours is 982 h (per year). The 'utopia point', i.e. the best theoretical performance (the simultaneous least discomfort and energy demand within the solution space) is 5900 kW h (168 kW h/m^2) and 580 h, which is also acquired as the learning target ($NLG = 1$).

Two (2) reference periods (t_{ref}) are investigated, namely, 30 min and 60 min, which represent the resolution levels of the HVAC system control. Salimi & Hammad (2020) point out that the cases with the resolution level of 60 min generate slightly better optimal solutions than those obtained from the cases with the application of local control with 30 min resolution. This is mainly due to larger energy savings that compensate the slight loss of comfort. The reason for this is the delay of the HVAC system control.

A modification of Eq. (12) yields the normalized learning gain, which (for the *i*-th performance indicator) can be written as follows:

$$NLG_i = \frac{P_{i,0}^{unc} - P_i^c}{P_{i,0}^{unc} - P_{i,\infty}} \tag{13}$$

where $P_{i,\infty}$ is the value of the *i*-th performance indicator at the learning target ('utopia point').

Applying Eq. (13) to the energy consumption of Case 4 (i.e. simple HVAC schedule with the 60-min resolution), the normalized learning gain is $NLG_{e,60\text{min}} = (6266–6138) / (6266–5900) = 0.35$.

Given that the learning method is two-objective optimization, the *NLG* is calculated as a single number by aggregating two performances, i.e. energy (subscript e) and (dis)comfort (c). The approach of Salimi and Hammad (2020) yields Pareto optimality, where a poor performance with respect to one objective may be compensated by good performance with respect to another. Hence, the indicator-specific achievements can be aggregated by a simple additive rule. Applying Eq.(7) for Case 4 and assuming equal weights for both performance indicators (i.e. $NLG_{e,60\text{min}} = 0.35$, $NLG_{c,60\text{min}} = 0.84$, $w_e = 0.5$, $w_c = 1 - w_e = 0.5$), the normalized learning gain is $NLG_{60\text{min}} = 0.5 \cdot 0.35 + 0.5 \cdot 0.84 = 0.59$. Again, the Average Learning Rates for the energy demand in Case 4, for example, $ALR_{e,60\ \text{min}} = (6266 - 6138)$ kWh/60 min = 2.1 kW h/min. For the discomfort hours, it is $ALR_{c,\ 60\ \text{min}} = (982 - 645)$ h/60 min =5.6 h/min.

The energy demands, discomfort hours and normalized learning gains are summarized case by case in Table 2. Here, the *NLG*s have been calculated as above, and the value $NLG = 0$ represents the baseline.

The data in Table 2 suggest that with the reference period of 30 min, the detailed schedule (Case 5, $NLG = 0.52$) slightly outranks the rival approach (Case 3, $NLG = 0.49 < 0.52$) in the sense of learning, whereas Case 4 ($NLG = 0.59$) becomes preferred to Case 6 ($NLG = 0.57 < 0.59$) at the period length of 60 min.

The graphs in Fig. 5 illustrate how the preferential treatment between the studied four cases (3,4,5,6) changes, when the weight of the energy consumption (w_e) (to be used in Eq. (7)) varies between zero and unity.

The graph in Fig. 5 indicates, for example, that when the ability of a control strategy to reduce energy consumption is weighted by 0.6 instead of 0.5, Case 6 becomes preferred in comparison with Case 4. Again, also Case 5 will outperform Case 4 if the weight is increased up to 0.8.

For the evaluation of the experience coefficient *X*, two main data sources are available in the work of Salimi & Hammad (2020), namely, 1) occupancy data and 2) simulation data. Again, both data sources can be classified into numerous subsets, but for the sake of simplicity only these main classes are included in the assessment.

The data resolution is very tight. For the occupancy data it is 1 s, whereas for the modeling and simulation it is 1 min. An acceptable baseline for a whole-building simulation study is 1 h. Hence, the normalized resolution for modeling and simulation is $Q_{r,s} = (1 - 1/60)$ h / (1 – 0) h = 0.98. For the occupancy data, the normalized resolution is $Q_{r,o} = (1 - 1/3600)$ h / (1 – 0) h ≈ 1.00.

The occupancy data set has been collected through monitoring realized occupancy, wherefore the normalized accuracy can be set to unity ($Q_{a,o} = 1$) for such applications, where the learning ability is evaluated for monitoring purposes. In assessments containing the need to predict the occupant behaviour, the accuracy should be set as $Q_{a,o} < 1$, for example, based on statistical uncertainty.

The other key quality attribute is simulation accuracy. According to Glasgo, Hendrickson, and Azevedo (2017), one of the major sources of errors in EnergyPlus-driven building simulations is occupancy, but in the present application, the impact of occupancy on the simulation error can be considered close to negligible since realistic occupancy data are used.

Yet some uncertainty related to the specification and modeling of the simulation case remains. Since Salimi & Hammad (2020) do not provide a comprehensive error analysis in their reporting, an estimate can be set, for example, on the basis of selected pre-defined standard requirements. Here, the suggested analysis method is Mean Bias Error (MBE), which de facto represents the mean difference between measured and simulated values (Coakley, Raftery, & Keane, 2014).

The ASHRAE Guideline 14 (ASHRAE Guideline 14-2014, Measurement of Energy, Demand, and Water Savings, ASHRAE, 2014) suggests that the monthly criterion for a calibrated model is MBE = 5 %. Inversely, assuming that this criterion is fulfilled, the normalized simulation accuracy would be $Q_{a,s} = 1 - 0.05 = 0.95$. In practice, the tolerance of 5 % is a tight criterion and commonly used for model calibration. For a more conservative estimate, a higher tolerance is suggested. Here, 20 % ($Q_{a,s} = 0.8$) is used, which is suggested by the International performance measurement & verification protocol IPMVP (EVO, 2007).

Table 2
Summary of energy demands, discomfort hours and learning gains.

Case	Description	Energy [kWh]	NLG_e	Discomfort hours [h]	NLG_c	NLG
2	Baseline	6266	0	982	0	0
3	Simple HVAC schedule, control resolution 30 min	6227	0.11	628	0.88	0.49
4	Simple HVAC schedule, control resolution 60 min	6138	0.35	645	0.84	0.59
5	Detailed HVAC schedule, control resolution 30 min	6116	0.41	728	0.63	0.52
6	Detailed HVAC schedule, control resolution 60 min	6076	0.52	729	0.63	0.57

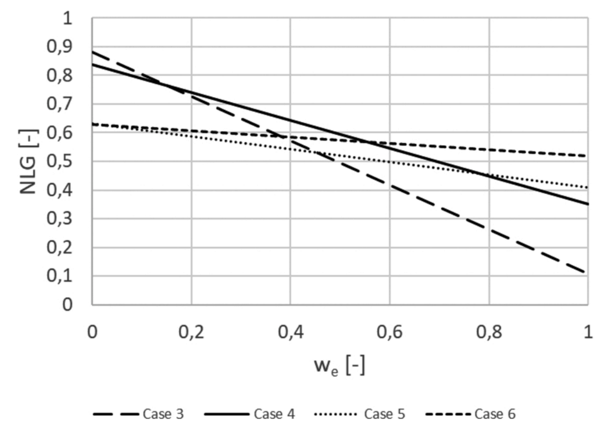

Fig. 5. Preferential treatment of the studied cases with weighted energy consumption.

Because the tolerance of the simulation includes a comparison between simulated and measured values, the measurement error basically should be considered, but here its impact is assumed negligible and it is omitted from the calculation of the *NLG* for the sake of simplicity.

Assume that both quality attributes (resolution, accuracy) are treated with equal weighting. The aggregated quality indicator for occupancy data (o) is $Q_o = w_r \cdot Q_{r,o} + w_a \cdot Q_{a,o} = 0.5 \cdot 1.0 + 0.5 \cdot 1.0 = 1.0$. Correspondingly, the aggregated quality indicator for simulation data (s) is $Q_s = w_r \cdot Q_{r,s} + w_a \cdot Q_{a,s} = 0.5 \cdot 0.98 + 0.5 \cdot 0.8 = 0.89$.

Since two data sources (occupancy and simulation) are included, the data diversity is also considered for the experience coefficient. To that end, assume that both data sets are in use and they are of equal value (diversity weight = 0.5). Thus, the experience coefficient $X = D_o \cdot Q_o + D_s \cdot Q_s = 0.5 \cdot 1.0 + 0.5 \cdot 0.89 = 0.95$. Here, the data diversity, i.e. the use of realistic occupancy data compensates the potential simulation errors related to model specification and modeling itself.

Again, the Learning Ability Index for Case 4 is $LAI_{60min} = 0.95 \cdot 0.59 = 0.56$.

3.3. Case study 3: Assessment of building's learning ability based on an occupant survey

Case study 3 has been selected with an aim to exemplify the calculation of the *LAI* on the basis of six (6) performance indicators and occupant survey data. The work of Karima and Altan (2017) has been chosen as the reference, since they conducted a questionnaire-based survey to get the perception of campus users (worker, facilitators, students, residents) about the intelligent performance of lighting, HVAC systems and intelligent building management system (IBMS) properties. The target building is an existing university campus building located in Abu Dhabi, United Arab Emirates. It belongs to the New York University Abu Dhabi and is located on a campus of 15.4 ha.

Karima and Altan (2017) mention that their study is based on the framework of the performance indicators for building intelligence, presented by Wong, Li, and Lai (2008). However, the survey actually asked the respondents' perception of the systems at a general level rather than that of the building's intelligence at the level of detail suggested by Wong et al. (2008).

The occupant survey included the respondents' perception of the buildings' intelligent performance with respect to the following items, namely:

1) Lighting System
2) Lighting Control
3) Natural Light
4) HVAC System
5) Temperature Control
6) Security System

The respondents evaluated the performance of all the six (6) items using a 4-grade scale (excellent, good, poor, very poor). To quantify the result, the numerical values from 1 to 4 are assigned to represent the verbal values from "very poor" (1) to "excellent" (4). Again, the overall value for each item is determined as the average of the grades given by individual respondents. For example, Karima and Altan (2017) state that six (6) of eight (8) respondents assigned the grade "good" (3) to lighting control, whereas two (2) respondents assigned the grade "poor" (2). Therefore, the intelligent performance of lighting control (index number 2) is defined as $P_2 = (6 \cdot 3 + 2 \cdot 2) / 8 = 2.75$.

It is notable that the grading scale in Karima and Altan (2017) does not provide a grade to represent a neutral opinion. Their survey only asks the respondents to evaluate the building's intelligence as either good or poor. In general, the well-known Likert scale would be more useful in surveys like this, since it encompasses the neutral grading option 'neither good nor poor'.

The survey presented by Karima and Altan (2017) can be interpreted as a pre-learning test, which reports the current level of the building's intelligence and establishes the baseline for calculating the *NLG*. The challenge is that there is a sparsity of building-related monitoring studies in the scientific literature, and hence a lack of post-testing data. Hence, Case study 3 remains at a hypothetical level, but yet the results as such can be considered useful to demonstrate the assessment of a building's learning ability on the basis of survey data.

Since no "post-test" results are available, a set of hypothetical monitoring (post-test) survey results are generated to simulate the building's learning during a reference period (t_{ref}), which in this type of an application may vary from weekly to annual level depending on the surveyor's preferences. To that end, assume that improvements for any or all the six (6) items will be made by 'soft' measures, such as training algorithms or optimization. Again, assume that the same respondents who took to the first questionnaire also participate in the monitoring survey and either keep their grade or improve it.

Without assuming anything else about the respondents' behaviour or performing any calculations, it can be stated that the building at least keeps its level of intelligence at the level of the baseline (i.e. $NLG = 0$). The 'utopia point' refers to a situation where all the respondents evaluate the performance of all the six items with the grade 'excellent' in the monitoring test, which results in the normalized learning gain $NLG = 1$. Including a further assumption that the survey result follows the normal (Gaussian) distribution, the aggregated *NLG* of around 0.5 has the highest probability.

To simulate the respondents' behaviour in one possible scenario, assume that they elevate their grades by one at maximum in comparison with the first survey (baseline). Again, all those who gave the grade 'excellent' in the first survey (baseline), keep their grading unchanged.

The upper part of Table 3 summarizes the number of respondents and their grades under the title Baseline (pre-learning) as provided by Karima and Altan (2017). The lower section of Table 3 presents the corresponding data as well as the calculated performances and *NLG*s for the hypothetical monitoring test. Concerning the lighting control, for example, five (5) of eight (8) respondents assign the grade 'excellent' (4) in the monitoring survey, whereas three (3) respondents assign the grade 'good' (3). In other words, one of the respondents who graded the lighting control as 'good', has kept the grading unchanged and all the rest have elevated it by one grade in comparison with the first survey. Now, the performance of lighting control (index number 2) is $P_2 = (5 \cdot 4 + 3 \cdot 3) / 8 = 3.63$. Again, the normalized learning gain for the lighting control is $NLG_2 = (3.63 – 2.75) / (4.00 – 2.75) = 0.70$.

The aggregated *NLG* (six items) for the hypothetical monitoring phase (using the data in Table 1 and assuming equal weighting between items) is $NLG = (0.17 + 0.70 + 0.20 + 0.40 + 0.00 + 0.57) / 6 = 0.34$. Again, it can be stated that if all the grades (apart from the 'excellent' ones) would be elevated by unity, the highest aggregated *NLG* would be 0.68 and the value $NLG = 0.68/2 = 0.34$ would have the highest probability in the normal distribution. Hence, the data in the lower section of Table 3 represent well the case that realizes with the highest probability.

A reasonable starting point for the assessment of quality and diversity of survey data for the calculation of the experience factor is its statistical and demographic representativeness. Here, the suggested performance indicators are the confidence level (i.e. the probability that the survey accurately reflects the attitudes of the population) and the margin of error (i.e. the range inside which the population's responses may deviate from those of the sample's). The population is limited to such persons who have an ability to assess the intelligent performance of the building through their own experience. Moreover, the quality of the survey data is affected by the respondents' level of knowledge and skills related to the systems to be assessed.

The limitation of the work of Karima and Altan (2017) is that only five (5) to eight (8) students of the student population (+1600 students according to the university's homepage) of the New York University Abu Dhabi responded to the survey. On the other hand, all the respondents were students, whereas the survey was originally targeted to four (4) groups of occupants in total (students, workers, residents, facilitators). The reported explanation for such a limited participation was that the survey was conducted during summer period.

To obtain the margin of error 5 % at the standard confidence level of 95 %, the sample size should be 310. The conclusion is that the results of

Table 3
Summary of data used for the calculation of normalized learning gains.

Baseline (pre-learning)	Excellent	Good	Poor	Very poor	*P*	*NLG*	Respondents
Lighting system	1	4	1	2	2.50	0.00	8
Lighting control	0	6	2	0	2.75	0.00	8
Natural light	3	5	0	0	3.38	0.00	8
HVAC system	0	2	5	1	2.13	0.00	8
Temperature control	0	0	0	5	1.00	0.00	5
Security system	2	5	1	0	3.13	0.00	8
Monitoring (post-learning)	**Excellent**	**Good**	**Poor**	**Very poor**	***P***	***NLG***	**Respondents**
Lighting system	2	3	2	1	2.75	0.17	8
Lighting control	5	3	0	0	3.63	0.70	8
Natural light	4	4	0	0	3.50	0.20	8
HVAC system	2	3	3	0	2.88	0.40	8
Temperature control	0	0	0	5	1.00	0.00	5
Security system	5	3	0	0	3.63	0.57	8

Karima and Altan (2017) simply do not provide sufficient data for a rigorous assessment of the experience coefficient in Case study 3. Therefore, the calculation of an experience coefficient from occupant survey data is demonstrated in hypothetical terms, assuming that the number of students is 1600 and that of faculty members is 400, which is in line with the data provided by the web page of the New York University Abu Dhabi. Thus, the population size is 2000 and it is assumed that all these individuals have an ability to assess the intelligent performance of the building on the basis of their personal experience and/or professional knowledge/skills.

Basically, the survey responses of the students and the faculty staff can be interpreted as two separate data sources. This strategy is useful, since the levels of experience and knowledge of these two respondent groups differ from each other in high probability. Here, the level of experience might be assessed within the range 0…1 so that, for example, the students represent the level 0.5 and the faculty members the level 1.0. These values could then be included into the quality coefficient Q. In real-life research, they should be acquired either from an external expert assessment or a separate survey for both respondent groups (self-evaluation).

The quality coefficient is also affected by the error of margin of the survey results. Assuming that 50 % of the entire student population (= 800 students) respond to the survey, the margin of error with the standard level of confidence (95 %) is 2 %. The statistical quality of the sample can be estimated as its complement, i.e. $Q = 1 - 0.02 = 0.98$.

Now, the aggregated quality coefficient of the survey data collected from students is $Q_s = 0.5 \cdot 0.98 = 0.49$ including both the level of experience and statistical data quality.

The diversity weight is simply the ratio of the sample size to the population size. The diversity weight is calculated from the total population (including all the respondent groups) and the sum of diversity weights over the sample domain is not necessarily unity (albeit unity at maximum). This approach makes it possible that both the participation activity and the expertise of all the respondent groups affect the experience factor X. In the case where the sample size of students is 800 of the total population of 2000, for example, the diversity weight is $D_s = 800/2000 = 0.4$. Considering that no responses from the faculty members are available, the diversity factor of the faculty members is $D_f = 0$. Correspondingly, the highest possible diversity factor for students in this population is 0.8, whereas for the faculty staff it is 0.2.

Implementing the aforementioned assumptions with the principles explained in Section 2, the experience factor for the sample of 800 (of 1600) students and 200 (of 400) faculty members is $X = D_s \cdot Q_s + D_f \cdot Q_f = 0.4 \cdot 0.49 + 0.1 \cdot 0.95 = 0.29$.

The value of *LAI* would be calculated using Eq. (8) as shown with details in Case studies 1 and 2. Therefore, it is not considered meaningful here to repeat the same calculation with numbers. On the other hand, completing the *LAI* calculation based on hypothetical monitoring data would not add value to Case study 3, either.

4. Discussion

In previous sections, the development of the Learning Ability Index (*LAI*) was explained and its implementation was demonstrated in three different case studies. It was shown that the *LAI* is useful when indicating the learning ability of a given process or whole system by way of a single, dimensionless number. The system's learning gain has to be tested in two points in time, i.e. initially (pre-learning) and in the monitoring phase (post-learning). The case studies show that the assessment based on the *LAI* is applicable to various sources of learning data and a variety of learning methods. Instead, the calculated *LAI*s only apply to the cases they represent. However, since the output is a single number between zero and unity, different learning strategies can be compared with other ones in proportional terms. Conversely, it is not possible to state on the basis of Case studies 1&2, for example, that the RF algorithm is preferred to the LM algorithm or that certain control scheme could be recommended for any building of the same type. Again, the Case study 3 is based on an occupant survey with a very limited number of respondents, and it does not encompass post-testing. Thus, the calculation results of Case study 3 are hypothetical. In the present study, learning has been defined as a change in the value of the intelligence function of the assessed process or system. Here, a system can be – at least in theory – equipped with an increased learning ability just by updating an application or installing a new one. Thus, learning ability is rather associated with the learning rate (effectiveness of learning) and the quality and diversity of the learning material than the amount of it. Correspondingly, the assessment of 'learning ability' differs from the assessment of just 'learning' in the sense that the effectiveness of learning (elapsed time) and the quality and diversity of learning material are considered. Again, the assessment contains an implicit assumption that the readiness for learning (learning capability) exists. Hence, the *LAI* is a performance indicator of different learning methods, materials, and strategies rather than an indicator of their presence.

Volkov (2013) and Batov (2015) included the idea of learning rate and latency (systemic delay) in their definition of Building Intelligence Quotient (BIQ). Since the *LAI* is virtually a quantitative measure of learning ability (it essentially includes the timely dimension and thus the learning rate) rather than just the description of the learning path (which may be long), it is also a measure of building intelligence.

The more processes there are observed and initiated by building AI, the better *LAI* depicts the building's intelligence. In the definition of BIQ by Volkov (2013) and, particularly, its explanation by Batov (2013), the BIQ ($\in [0,1]$) can be interpreted as the ratio of the sum of AI-initiated set of processes to the entire set of processes controlling the automatically observed operational parameters (including those controlled by humans or non-intelligent automation systems). Moreover, the sum over these processes is weighted by their significance for the system control. For a very simple example, if space heating and the heating of supply air are considered separate processes with equal weights, the significance of both processes is set to unity and space heating is entirely initiated by AI, the intelligence quotient of the temperature control is $BIQ = 1/(1 + 1) = 0.5$.

Actually, the BIQ model is not that simple, but the challenge is to obtain descriptive enough significances for each process. Referring to the requirement presented by Batov (2015) to consider the accuracy or a control process in tandem with its learning speed and latency, one option would be to use a process-specific *LAI* to indicate its significance. Assume, for example, that the *LAI* of an AI-initiated process is 0.9, whereas that for a 'non-intelligent' process controlling the same parameter is 0.7. Now, $BIQ = 0.9/(0.9 + 0.7) = 0.56$. Here, it is notable that when the significance of AI-initiated processes becomes closer to unity, the *BIQ* also approaches unity, wherefore information about the accuracy, learning speed and latency will be lost. Therefore, it is reasonable to state that the *LAI* as such would be an appropriate indicator for building intelligence since the learning rate as a measure of intelligence always will be considered.

In the suggested assessment method, the learning target is defined implicitly, i.e. the reference time is set instead of a quantified learning outcome (within the reference time). Inversely, the preferred level is always 100 % of the implicit learning target (i.e. $NLG = 1$). Defining the implicit learning target as well as setting up the reference time is the evaluator's task. To that end, a guideline can be given that the learning target should represent a 'utopia point', which is the best theoretically available intelligent performance within the given framework of performance indicators. Hence, the learning target is not to be just implicit, but also generic enough. On the other hand, the value of the *LAI* is always case-specific and depends on the starting level (baseline), wherefore the only (and sufficient) condition for a 'good' performance is that the *LAI* should be close to unity.

The *LAI* is recommended to be used retroactively for monitoring purposes, which means that the assessment of the building intelligence always includes the analysis of the present status (pre-learning), setting up the length of the reference period and the baseline, and finally, the determination of the building intelligence after the reference period via post-testing (monitoring) in a similar way as the students' learning is assessed in the educational settings. Hence, the assessment also relies on realistic performance data and the correcting actions may be conducted reactively.

Yet the *LAI* can be used in predictive or proactive settings, for example, with an aim to find the most preferred control strategies via optimization. Here, the challenge is that the target system and proposed learning strategies have to be simulated to predict the post-test performance, which may be computationally expensive. The use of the *LAI* for the simultaneous comparison of several learning strategies is a challenge, since setting up the baseline is not necessarily unambiguous.

The *LAI* is recommended for assessing the learning impacts related to software, systemic or organizational learning. However, there is no limitation to use the method to evaluate improvements due to upgrading hardware, either, since the *LAI* is basically a generic indicator of development. Interesting avenues for the future research would be studying the applicability of the *LAI* to evaluate human-building learning interactions through rigorous pre-post testing, to assess buildings' ability to correct erroneous functionality and to investigate the potential of digital twins (DT) in enhancing buildings' learning ability.

Selected strengths and weaknesses of the *LAI* have been summarized in Table 4.

Table 4
Summary of strengths and weaknesses of the Learning Ability Index (*LAI*).

Strengths (+)	Weaknesses (−)
+ Yields a single, dimensionless number to indicate the building's learning ability	− Requires two data recordings/tests (pre-test for the baseline, post-test for monitoring)
+ Provides an explicit and illustrative numerical value between 0…1 (or 0 %…100 %) to indicate "good" (close to unity) and "poor" (close to zero) performance, which allows the generic comparison of various learning strategies in proportional terms	− Relies on case-specific choices including the baseline criteria, the learning target ('utopia' level), the reference time and possibly weighting between KPIs and/or quality indicators → The result is somewhat subjective and it cannot be generalized as such to buildings other than the target building (e.g. smart buildings of a similar type).
+ Is applicable together with the diversity of quantitative/qualitative KPIs/ definitions of smartness/intelligence and types of AI	
+ Flexible in terms of the reference time (from seconds to the whole building life cycle)	
+ Can be generalized to development index or sustainability index	
+ Utilizes data acquired from various sources including real-time measurements, simulation results and occupant surveys	
+Easy to implement using computer algorithms	
+ Computationally simple and affordable	

5. Conclusions

This paper describes the assessment of the learning ability of buildings using a novel metrics, i.e. Learning Ability Index (*LAI*), which is a single, dimensionless number located between zero and unity. The method encompasses not just the assessment of the learning gain, but it also addresses the impacts of elapsed time plus the quality and diversity of learning material.

Particularly, the *LAI* is useful for monitoring the learning of building-integrated, AI-initiated processes with an aim to identify the preferred strategies to elevate buildings' intelligence. The method can be generalized to the evaluation of learning in a more extensive context (e.g. organizational learning, smart workspace) in terms of strategic optimization. It is applicable to not just the assessment of buildings and communities, but any system with AI-initiated unit processes, such as autonomous vehicles.

In the work presented in this article, the *LAI* was implemented and demonstrated in three different case studies chosen from the literature. The method was experienced flexible with respect to system boundaries (single control process, building, community) and the framework of key performance indicators (KPIs). The method is computationally affordable and easy to use. The *LAI* is also applicable as a generic indicator of system evolution or the indicator of process-specific learning speed, latency and accuracy when calculating Building Intelligence Quotient (BIQ). The challenge is that data have to be collected at two points in time (pre-learning and post-learning) and the time between these has to be fixed by the evaluator with case-specific justifications. Again, the evaluator has to determine an implicit learning target ('utopia point'), which is also case-specific.

There are several relevant topics for the future research. First, scientifically sound studies are needed to produce information and data on how the implementation of AI and ML may support shared learning between buildings and occupants. Second, there is a need of studies on how ICT-assisted learning methods (e.g. the use of digital twins (DT) to enhance the quality of learning data) may help in improving the building's learning ability. Third, there is a need to seek the pathway of the simplest available solutions to make the building learn as a whole (instead of training or enhancing single processes). The *LAI* is a potentially useful research tool for all these problems.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Al Dakheel, J., Del Pero, C., Aste, N., & Leonforte, F. (2020). Smart buildings features and key performance indicators: A review. *Sustainable Cities and Society, 61*, 102328.
Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology, 22*(1), 3–21.
Araya, D. B., Grolinger, K., El Yamany, H. F., Capretz, M. A. M., & Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings, 144*, 191–206.
Azuatalam, D., Lee, W.-L., de Nijs, F., & Liebman, A. (2020). Reinforcement learning for whole-building HVAC control and demand response. *Energy and AI, 2*, 100020.
Batov, E. I. (2015). The distinctive features of "Smart" buildings. *Procedia Engineering, 111*, 103–107.
Candanedo, L. M., Feldheim, V., & Deramaix, D. (2018). Reconstruction of the indoor temperature dataset of a house using data driven models for performance evaluation. *Building and Environment, 138*, 250–261.
Cao, L., Li, Y., Zhang, J., Jiang, Y., Han, Y., & Wei, J. (2020). Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Reports, 6*, 2751–2767.
Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews, 37*, 123141.
EPBD Recast. (2010). Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast). *Official Journal of the European Union* (2010) 18/06/2010.
EVO. (2007). *International performance measurement & verification protocol*. Efficiency Valuation Organisation.
Glasgo, B., Hendrickson, C., & Azevedo, I. L. (2017). Assessing the value of information in residential building simulation: Comparing simulated and actual building loads at the circuit level. *Applied Energy, 203*, 348–363.
Han, M., May, R., Zhang, X., & Jin, Y. (2020). A novel reinforcement learning method for improving occupant comfort via window opening and closing. *Sustainable Cities and Society, 61*, 102247.
Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2014). User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. *Energy and Buildings, 70*, 398–410.

Karima, M., & Altan, H. (2017). Interactive building environments: A case study university building in UAE. *Procedia Engineering, 180*, 1355–1362.

Karpook, D. (2017). *Question of the week: How buildings learn*. Retrieved from https://facilityexecutive.com/2017/12/how-buildings-learn-machine-learning/. Accessed December 7, 2020.

Konstantakopoulos, I. C., Barkan, A. R., He, S., Veeravalli, T., Liu, H., & Spanos, C. (2019). A deep learning and gamification approach to improving human-building interaction and energy efficiency in smart infrastructure. *Applied Energy, 237*, 810–821.

Lê, Q., Nguyen, H. B., & Barnett, T. (2012). Smart homes for older people: Positive aging in a digital world. *Future Internet, 4*(2), 607–617.

Li, D., Zhou, Y., Hu, G., & Spanos, C. J. (2016). Fault detection and diagnosis for building cooling system with a tree-structured learning method. *Energy and Buildings, 127*, 540–551.

Marinakis, V., Karakosta, C., Doukas, H., Androulaki, S., & Psarras, J. (2013). A building automation and control tool for remote and real time monitoring of energy consumption. *Sustainable Cities and Society, 6*, 11–15.

McGrath, C., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). *Learning gain in higher education*. Santa Monica, California and Cambridge UK: RAND. Retrieved fromhttps://www.rand.org/content/dam/rand/pubs/research_reports/RR900/RR996/RAND_RR996.pdfAccessed December 7, 2020.

Mofidi, F., & Akbari, H. (2020). Intelligent buildings: An overview. *Energy and Buildings, 223*, 110192.

Piech, C., Bumbacher, E., & Davis, R. (2020). Measuring ability-to-learn using parametric learning-gain functions. *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)*, 496–502. Virtual conference. Retrieved from: https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_226.pdf. Accessed December 7, 2020.

Salimi, S., & Hammad, A. (2020). Optimizing energy consumption and occupants comfort in open-plan offices using local control based on occupancy dynamic data. *Building and Environment, 176*, 106818.

SCIS. (2017). *Monitoring KPI guide D23.1*. Retrieved from: https://smartcities-infosystem.eu/sites/www.smartcities-infosystem.eu/files/document/scis-monitoring_kpi_guide-november_2018.pdf. Accessed December 7, 2020.

Verbeke, S., Ma, Y., Van Tichelen, P., Bogaert, S., Waide, P., Uslar, M., et al. (2017). *Support for setting up a Smart Readiness Indicator for buildings and related impact assessment. Interim report*. Retrieved from: https://smartreadinessindicator.eu/sites/smartreadinessindicator.eu/files/sri_for_buildings_interim_report_20171212.pdf. Accessed December 7, 2020.

Vigna, I., Pernetti, R., Pernigotto, G., & Gasparella, A. (2020). Analysis of the building smart readiness Indicator calculation: A comparative case-study with two panels of experts. *Energies, 13*, 2796.

Volkov, A. A. (2013). Building intelligence quotient: Mathematical description. *Applied mechanics and materials* (pp. 392–395). Switzerland: Trans Tech Publications, 409–410.

Wang, B., Tao, F., Fang, X., Liu, C., Liu, Y., & Freiheit, T. (2020). Smart manufacturing and intelligent manufacturing: A comparative review. *Engineering*. https://doi.org/10.1016/j.eng.2020.07.017. In Press.

Wong, J., Li, H., & Lai, J. (2008). Evaluating the system intelligence of the intelligent building systems. *Automation in Construction, 17*(3), 303–321.

Xu, J., Lu, W., Xue, F., & Chen, K. (2019). Cognitive facility management: Definition, system architecture, and example scenario. *Automation in Construction, 107*, Article 102922.