



Measuring Sustainable Development Using an Extended Hellwig Method: A Case Study of Education

Ewa Roszkowska¹ · Marzena Filipowicz-Chomko²

Accepted: 6 September 2020 / Published online: 22 September 2020
© The Author(s) 2020

Abstract

Benchmarking the analysis of countries' performance in terms of sustainable development helps to understand the success factors of countries that over perform and to target priority issues of others with lower performance. However, assessing sustainable development comes with methodological challenges, including indicator standardization, aggregation and weighting. Our study significantly contributes to the measure of sustainable development by providing a new approach based on an extended Hellwig method. After describing the main limitations of existing methodologies, this paper's aim is twofold. First, we show that the proposed analytical framework allows for comparing the sustainable performance of EU countries on the national level. The extended Hellwig method takes into consideration EU targets and/or national targets in building patterns of development. Second, this framework is tested as a part of the evaluation of the implementation of the Europe 2020 strategy in the education area. The results obtained using the extended Hellwig method were compared with those obtained by means of the Education Index, TOPSIS and Ward technique. Our analysis showed the significant disparities in the implementation of the Europe 2020 strategy recommendations in the education area in 2015.

Keywords Sustainable development · Europe 2020 strategy · Education · Hellwig method · Ward method

Education is the most powerful weapon you can use to change the world.
(Nelson Mandela)

✉ Marzena Filipowicz-Chomko
m.filipowicz@pb.edu.pl

Ewa Roszkowska
e.roszkowska@uwb.edu.pl

¹ University of Białystok, Warszawska 53, 15-063 Białystok, Poland

² Białystok University of Technology, Wiejska 45A, 15-351 Białystok, Poland

1 Introduction

The Europe 2020 strategy (European Commission 2010) is the EU's agenda for growth and job creation, which emphasizes mutually reinforcing priorities of smart, sustainable and inclusive growth. For each of these priorities, the strategy defined five headline targets at the EU level, which belong to thematic areas covering issues of employment, education, poverty and social exclusion, climate change and energy, R&D and innovation. These targets should be met by 2020.

Composite indices are widely used as synthetic measures for ranking and benchmarking alternatives across complex concepts. Additionally, Multi-Criteria Decision Analysis (MCDA) offers flexible tools for the quantitative assessment and ranking of options according to multiple system dimensions, which are not directly measurable or clearly defined, such as sustainability, human development, competitiveness and quality of governance (Wang et al. 2009; Huang et al. 2011; Cinelli et al. 2014; El Gibari et al. 2018).

For building composite indices in transparent and representative manner as well as to avoid misleading results, the Organization for Economic Cooperation and Development (OECD) together with the Joint Research Centre (JRC) of the European Commission have published a "Handbook on Constructing Composite Indicators: Methodology and User Guide", which has become the reference for practitioners constructing composite indicators in a variety of fields. They have developed a ten-step methodological framework for constructing a composite indicator (OECD 2008). The steps are as follows: theoretical framework, data selection, imputation of missing data, multivariate analysis, normalization, weighting and aggregation, uncertainty and sensitivity analysis, back to the data, links to other indicators and the visualization of results.

Bandura (2011) provided a comprehensive inventory of over 400 country-level indices monitoring complex phenomena from economic progress to educational quality. Additionally, in the reviewed paper, Greco et al. (2019) identified a very high growth of composite indicators over the past 20 years, highlighting their popularity and applicability to several domains. Saltelli et al. (2011) pointed out some problems connected with statistical information in monitoring the progress of European countries towards the implementation of the Europe 2020 strategy. In the literature, we can find many propositions and studies contributing to the measurement of sustainable development (Colak and Ege 2013; Pasiemi 2012, 2013; Rappai 2016; Walheer 2018, among others). In the context of the 2030 Agenda for Sustainable Development (United Nations 2015), the question about synthetic indices, which might be used for the measurement of sustainable development taking into consideration EU targets and/or national targets, seems to be vital. This was the main motivation of our paper.

The aim of this paper is twofold. The first goal is to propose an analytical framework for comparing the sustainable performance of EU countries on the national level. To this aim, an extended Hellwig procedure is proposed, which allows us to take into consideration EU targets and/or national targets in building patterns of development. The second goal of this paper is to apply the proposed framework (the extended Hellwig method) to test its relevance and effectiveness in assessing the education component of the EU's agenda.

The multiple-criteria analysis of the implementation level of the Europe 2020 strategy recommendations is provided, and the spatial diversity of EU countries in the education area in 2015 is presented. A ranking of 28 EU countries was performed to understand the specificity of education in the context of sustainable growth with the application of a classical as well as extended Hellwig method. First, we show that the set of SDG 4 indicators

used in Hellwig procedure describes the phenomenon of education from the perspective of sustainable development better than indicators used in the Education Index. Hellwig measure is based on the process and quality of knowledge and employment rates of recent graduates, while the Education Index is based only on years of schooling. Second, a sensitivity analysis is performed. We show that the results obtained by Hellwig procedure with those obtained by the extended TOPSIS technique (Roszkowska and Filipowicz-Chomko 2019) with the same set of indicators are consistent. Third, we also identify the group of EU countries that are similar to one another but different from other groups of countries based on the studied indicators. For this purpose, cluster analysis, with application of Ward method, was used. The EU countries were divided into clusters of regions grouping the most similar countries due to the level of education, using the hierarchical Ward agglomerative procedure. Hellwig procedure allows us to classify the countries into four groups characterized by different levels of development (very high, medium-high, medium-low, very low). The classification results obtained by Hellwig and Ward methods were also compared. A sensitivity analysis for grouping was performed.

The paper is organized as follows. Section 2 presents a short literature overview on sustainable development. Section 3 presents the taxonomic Hellwig method and a new extended Hellwig measure and describes the Education Index and Ward method. In Sect 4, the goals in the education area, the empirical data and the obtained results are described. Section 5 compares the research findings on the rank ordering of EU countries obtained by the proposed techniques and by other methods. The summary concludes the paper.

2 Literature Review

The European Commission proposed several indicators to monitor the Europe 2020 strategy: Tertiary education attainment (TEDU), gross domestic expenditure on R&D (GERD), greenhouse gas emissions, the share of renewable energy in gross final energy consumption (RNEW), the energy intensity of the economy (EINT), the employment rate of the population aged 20–64 (EMPL) and early leavers from education (SCHO) (OECD 2008). To know how far each individual state is from reaching the desired objectives, the composite indicators have been defined. In addition to those recommended by the OECD, other composite indices have independently been suggested in this thematic area by researchers (see Colak and Ege 2013; Pasimeni 2012, 2013; Bolcárová and Kološta 2015; Rappai 2016; Walheer 2018 among others). Colak and Edge (2013) assessed the EU's new growth strategy from a different perspective by evaluating the performances of the EU member and candidate countries with respect to the three growth priorities and the overall strategy by constructing a composite index inspired by the HDI. Pasimeni (2012, 2013) for the purpose of cross-country comparison, selected the geometric method of the index, i.e., including a mechanism to reward more balanced development profiles. The smart (SMGI), sustainable (SUGI) and inclusive (INGI) growth indices were aimed at measuring three dimensions of the Europe 2020 strategy. Bolcárová and Kološta (2015) created an aggregated index of sustainable development (SD) from the EU set of SD indicators for each EU 27 country based on principal component analysis (PCA) and correlated this new aggregated index with economic growth Rappai (2016). Proposed a new complex index based on self-weighting average and Mahalanobis distance, which measures how close the member states are to the completion of the strategy, and considered the heterogeneity of growth paces across countries. Walheer (2018) suggested a decomposition of the composite index

by distinguishing between three different components: country-, group-, and objective-specific indices. Walheer (2018) also described “A country-specific index showing how each country performs with respect to the best performer for each year, a group-specific index that indicates how the group performs for every year, and an objective-specific index that shows whether, in principle, the targets are reachable for the period”.

In recent years, the methodology of multiple analysis has developed significantly, and some interesting complex approaches have been proposed to eliminate the weaknesses of composites (Maggino 2017; Greco et al. 2019). Maggino (2017) considered some important methodological aspects and issues that should be taken into consideration when measuring complex social phenomena. Mazziotta and Pareto (2017) highlighted that building a composite index is a delicate task. They pointed out the importance of the following steps: selecting a group of individual indicators, normalizing, standardization indicators and validating the composite index. Greco et al. (2019) offered an outlook on the advances made in the methodological approaches concerning composite indicators. They focused, in particular, on two steps that were criticized the most: weighting and aggregation. In addition, they reviewed the robustness analysis.

Sustainable development assessment is a multi-dimensional problem comprising economic, social and environmental issues (United Nations 1992, 2015; Rio Declaration 1992). Janeiro and Patel (2015) noticed that sustainability assessment is a Multi-Criteria Decision Making (MCDM) problem since it looks for compromises among conflicting criteria and indicators. There is a wide literature on implementation sustainability using multi-criteria decision methods. To build a composite indicator, several aggregation procedures have been used. A good overview may be found in Azapagic and Perdan (2005a, b); Munda (2005a, b); Singh et al. (2009); Rowley et al. (2012); Cinelli et al. (2014); Diaz-Balteiro et al. (2017); Gana et al. (2017), Büyüközkan and Karabulut (2018); and Ture et al. (2018), among others.

Azapagic and Perdan (2005a, b) discussed three steps: problem structuring, problem analysis and sustainability assessment and guidance on the choice of the most appropriate MCDA method to measure sustainability. Munda (2005a, b) described compensatory and non-compensatory aggregation methods from the weak and strong sustainability perspectives. Singh et al. (2009) provided an overview of various sustainability indices used in policy practice. They measured sustainable development taking into consideration scaling, normalization, weighting and aggregation methodologies. Rowley et al. (2012) discussed the theoretical implications of sustainability analysts and environmental decision makers' choice of the MCDA method in a given problem context. Cinelli et al. (2014) described five MCDA methods (i.e., MAUT (*Multiple Attribute Utility Theory*), AHP (*Analytic Hierarchy Process*), PROMETHEE (*Preference Ranking Organization METHOD for Enrichment of Evaluations*), ELECTRE (*ELimination and Choice Expressing REality*) and DRSA (*Dominance-based rough set approach*) with respect to some criteria that sustainability assessment tools should satisfy, among which are a life cycle perspective, thresholds and uncertainty management, software support and ease of use. Diaz-Balteiro et al. (2017), based on the set of 271 papers appearing in the ISI Web of Science database, studied and critically assessed the literature published on these topics. The results show that these techniques have been applied to a great variety of problems, levels, and sectors related to sustainability. Gana et al. (2017) discussed the advantages and disadvantages of nine weighting methods and three aggregation methods. They also proposed a procedure for choosing the most suitable weighting and aggregation methods based on four steps: Research purposes, spatial and temporal scales, and sustainability perspectives. They noticed that several sustainability indices have been built using an equal weighting strategy, when all the

indicators are considered equally important or no statistical or empirical evidence supports a different scheme. Büyüközkan and Karabulut (2018) reviewed the state of sustainability performance evaluation (SPE) literature with a business perspective to capture needs and develop research recommendations. They identified 128 articles between 2007 and 2018. The obtained results suggested that sustainability performance evaluation models should be more balanced, the suitable criteria and their interrelations should be well defined, and the subjectivity of qualitative criteria inherent to sustainability indicators should be considered. Ture et al. (2018) measured, monitored and compared the countries according to their performance by using the VIKOR (*Vlse Kriterijumska Optimizacija Kompromisno Resenje*) and the TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*) methods. The authors highlighted that the proposed methods can be applied to assess countries in terms of the EU 2020 strategy, which has multi-dimensionality targets.

It is worth noting that each method used by researchers has its advantages as well as limitations, assumptions and characteristics. On the other hand, all of them allow integrating the set of indicators and are capable of considering such a broad spectrum of criteria, including various economic, financial, demographic, educational and innovational areas. All the aforementioned methods, excluding the extended TOPSIS (Roszkowska, Filipowicz-Chomko 2019), do not provide the possibility of including EU targets and/or national targets in their procedure. The extended Hellwig procedure differs from the extended TOPSIS (Roszkowska, Filipowicz-Chomko 2019) since it measures only the distances to the pattern of development. The extended TOPSIS method takes into consideration EU targets and/or national targets in building PIS (*positive ideal solution*) and NIS (*negative ideal solution*), and then, the distances to ideal and anti-ideal patterns are calculated. In this paper, another approach is proposed based on Hellwig procedure. Additionally, those two approaches were compared.

3 Methodology

The Taxonomic Measure of Development (TMD) introduced by Hellwig in 1968 (Hellwig 1968) is a method that allows the measurement of sustainable development. In our paper, we present the classical approach to this method, and then, we propose a modification to this procedure. The TMD allows us to rank objects (for example, countries) from the worst to the best based on the level of development of the phenomenon under evaluation. For each aspect of the analysed phenomenon, the set of diagnostic variables (indicators) is selected based on substantive and statistical reasons, and relevant statistical data are gathered; next, a synthetic measure (TMD) is calculated. It determines the Euclidean distance of each object from the reference object, the so-called development pattern. Hellwig technique is similar to the TOPSIS procedure, which is often applied in MCDM (Hwang and Yoon 1981). The TOPSIS procedure uses the concept of distances to the ideal and negative ideal solutions, while Hellwig method uses only the concept of distance to ideal solution (pattern development). The basic concept of TOPSIS is that the chosen alternative should have the shortest distance to the ideal solution and the farthest distance to the negative ideal solution. However, contrary to the TOPSIS procedure, Hellwig method uses only the concept of an ideal solution (pattern development). Hellwig method was originally proposed in 1968, long before TOPSIS, as a taxonomic method for international comparisons of economic development of countries (Hellwig 1968). It was disseminated in the international literature in 1972 with the realization of the UNESCO research project on the human

resources indicators for less developed countries (Hellwig 1972a, b). Hellwig method was presented in UNESCO (1972a), where its application to a group of countries with respect to health development was described. UNESCO (1972b) has been used in development areas such as health development, human resource development and educational development. Other mentions and applications of Hellwig method can be found in Baster (1972), Di Domizio (2008), Pawlas (2016), Bilgin et al. (2015), and Reiff et al. (2016), among others.

3.1 The Classical Hellwig Procedure

Let $A = \{A_1, A_2, \dots, A_n\}$ be the set of objects (e.g., countries) connected with the examined phenomenon, $X = \{C1, C2, \dots, Ck\} = P \cup N$ – the set of individual indicators (e.g., sustainable development (SD) indicators from Eurostat) characterizing the phenomenon, where P and N are the sets of positive and negative indicators, respectively. The classical Hellwig procedure consists of the following steps:

Step 1 Defining the evaluation vector for each object: $A_i = [x_{i1}, x_{i2}, \dots, x_{ik}]$ where x_{ij} is the value of the j -th indicator for the i -th object, $i = 1, \dots, n, j = 1, \dots, k$.

Step 2 Defining the vector of weights $w = [w_1, \dots, w_k]$, where $w_j > 0 (j = 1, \dots, k)$ is the weight of indicator C_j and $\sum_{j=1}^k w_j = 1$.

Step 3 Building the pattern of development as an abstract point $P_0 = [x_1^+, x_2^+, \dots, x_k^+]$ in accordance with Hellwig concept (Definition 6, Hellwig 1968) defined as

$$x_j^+ = \begin{cases} \max_i x_{ij} & \text{if } x_{ij} \in P \\ \min_i x_{ij} & \text{if } x_{ij} \in N \end{cases}$$

Step 4 Building the normalized vector $\bar{A}_i = [\bar{x}_{i1}, \dots, \bar{x}_{ik}]$ for the i -th object using the standardization formula:

$$\bar{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j},$$

where $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$.

Step 5 Building the weighted normalized i -th vector for each object $\tilde{A}_i = [\tilde{x}_{i1}, \dots, \tilde{x}_{ik}]$ where

$$\tilde{x}_{ij} = w_j \bar{x}_{ij}.$$

Step 6 Calculating the distances (d_{i0}) of the i -th object from the object pattern (A_i, P_0) by using the classical Euclidean distance measure:

$$d_{i0} = \sqrt{\sum_{j=1}^k (\tilde{x}_{ij} - w_j)^2}.$$

Step 7 Calculating the synthetic measures of development. For the i -th object, we compute

$$H_i = 1 - \frac{d_{i0}}{d_o}.$$

where $d_0 = \bar{d} + 2S$, for $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_{i0}, S = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d})^2}$.

Step 8 Ranking of the objects according to descending H_i .

H —determined in accordance with the described procedure is a normalized measure, which in most cases has values from 0 to 1. The higher the synthetic measure value is, the higher the position of a given object in the ranking.

3.2 The Extended Hellwig Procedure

In the classical Hellwig procedure, the abstract-point P_0 is defined by means of the maximum and minimum relations. This point is common for all objects (countries). In the modified method, we departed from Hellwig approach. To take into consideration EU targets and member states' individual targets, we introduced the notion of an individual pattern of development $P_{i0} = [x_{i1}^+, x_{i2}^+, \dots, x_{ik}^+]$ for the i -th object separately with target values (benchmarks) as coordinates. More precisely, in our approach, x_{ij}^+ is the value of the EU/national target of the j -th indicator for the i -th country.

The main change concerns the normalization formula (Step 4). We must consider two situations for positive and negative indicators:

1) For positive indicators,

- (a) if the target is reached, i.e., $x_j^- < x_{ij}^+ \leq x_{ij}$ then $\frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} \geq 1$;
- (b) if the target is not reached, i.e., $x_j^- < x_{ij} < x_{ij}^+$ then $\frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} < 1$;

(2) For negative indicators,

- (a) if the target is reached, i.e., $x_{ij} \leq x_{ij}^+ < x_j^-$ then $\frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} \geq 1$;
- (b) if the target is not reached, i.e., $x_{ij}^+ \leq x_{ij} < x_j^-$ then $\frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} < 1$.

As a consequence, we obtain modified stages of the classical procedure:

Step 4a Building the normalized vector $\bar{A}_i = [\bar{x}_{i1}, \dots, \bar{x}_{i8}]$ for the i -th object (country) by using the formula

$$\bar{x}_{ij} = \begin{cases} 1 & \text{if } \frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} \geq 1 \\ \frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} & \text{if } \frac{x_{ij} - x_j^-}{x_{ij}^+ - x_j^-} < 1 \end{cases}$$

where x_{ij}^+ is the development pattern of the i -th object, which is individually defined (for instance, by means of assumed EU or national targets) and $x_j^- = \begin{cases} \max_i x_{ij} & \text{if } x_{ij} \in N \\ \min_i x_{ij} & \text{if } x_{ij} \in P \end{cases}$

Step 5a Building the modified weighted normalized i -th vector $(\tilde{A}_i)_m = [(\tilde{x}_{i1})_m, \dots, (\tilde{x}_{ik})_m]$ for each object (country) where

$$(\tilde{x}_{ij})_m = w_j \bar{x}_{ij}.$$

Step 6a Calculating the distances $(d_{i0})_m$ of the i -th object (country) from the object pattern (A_i, P_{i0}) by using the classical Euclidean distance measure:

$$(d_{i0})_m = \sqrt{\sum_{j=1}^k \left((\tilde{x}_{ij})_m - w_j \right)^2}.$$

Step 7a Calculating the modified synthetic measures of development for the i -th object (country):

$$(H_i)_m = 1 - \frac{(d_{i0})_m}{(d_0)_m}$$

where $(d_0)_m = \bar{d} + 2S$, for $\bar{d} = \frac{1}{n} \sum_{i=1}^n (d_{i0})_m$, $S = \sqrt{\frac{1}{n} \sum_{i=1}^n ((d_{i0})_m - \bar{d})^2}$.

Step 8a Ranking of all objects (EU countries) according to descending $(H_i)_m$.

Moreover, taking into account the values of synthetic measure H , objects (countries) were grouped. The following division, which allowed us to distinguish the objects characterized by different levels of development, was used. The objects were classified into the following groups:

Group 1: $\hat{H} \in [\bar{H} + S_H; \max\{H_i\}]$ represents a very high level of sustainable development,

Group 2: $\hat{H} \in [\bar{H}; \bar{H} + S_H)$ represents a medium–high level of sustainable development,

Group 3: $\hat{H} \in [\bar{H} - S_H; \bar{H})$ represents a medium–low level of sustainable development,

Group 4: $\hat{H} \in [\min\{H_i\}; \bar{H} - S_H)$ represents a very low level of sustainable development, where $\hat{H} \in \{H1, H2, H3\}$, \bar{H} and S_H are the average value and the standard deviation of synthetic measures, respectively

3.3 Framework Measuring Sustainable Development Based on Hellwig Approach

The proposed framework measuring sustainable development based on Hellwig method consists of seven main stages, as presented in Fig. 1.

The first four steps are similar to those in other synthetic measures. The first step is the problem description, the second requires determining the set of evaluated objects, while the third identifies individual indicators. Eurostat collected 232 EU Sustainable Development Goals (SDGs) indicator set, which constituted the basis for assessing the level of implementation of Europe 2020 with respect to 17 goals as well as the main 169 EU targets identified in this strategy and national targets accepted by the EU member states. Guided by these aspirational and EU targets, each government is also expected to pursue its own national targets, taking into account national circumstances. Indicators are required for monitoring the level of achievement for these targets.

The fourth step in the process involves assigning weights to the indicators according to their importance. The literature provides various approaches to weight the indicators (Gana et al. 2017; Roszkowska 2013). Maggino and Ruvigliani (2009) noted that equal weights were used in most applications and said that a reason why this happens is that the theoretical structure attributes each indicator the same adequacy in defining the variable to be measured.

In the fifth step, the pattern of development, abstract or individual based on the benchmark values, is built. In the next step, Hellwig algorithm is employed to evaluate and rank the order of the objects. Indicators were converted to a common scale by the process of standardization based on the mean and standard deviation of indicators. Next, the

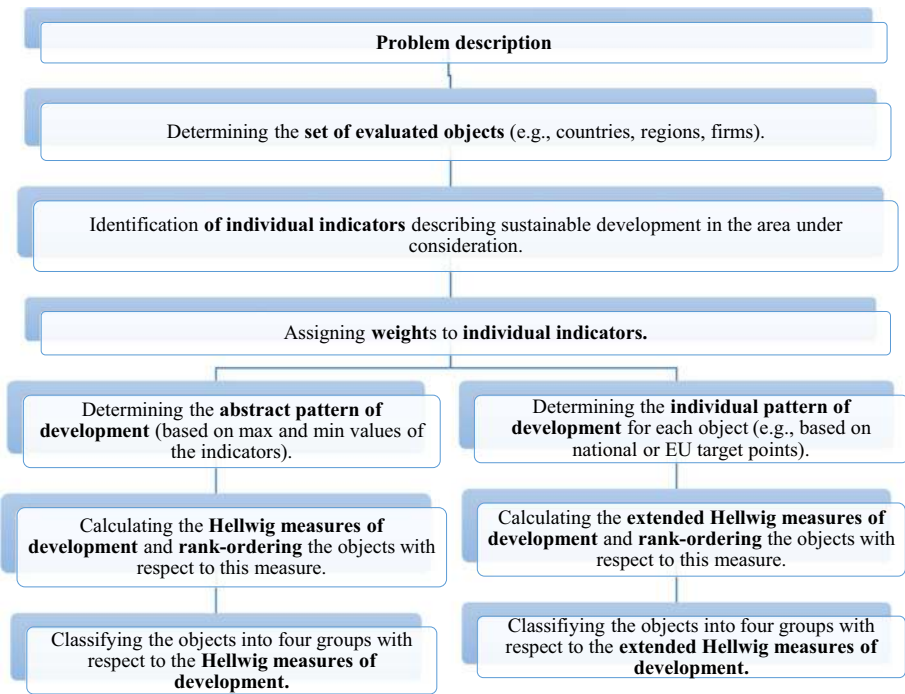


Fig. 1 Measuring sustainable development based on Hellwig methods

numerical distances to the pattern of development are built. Finally, the objects are grouped into four classes according to the values of the synthetic measure.

From this perspective, the main advantage of the extended Hellwig approach is that it allows us to evaluate EU and national targets when measuring sustainable development. However, there are two disadvantages: One is that some attempts have been made to reduce the number of indicators through correlation analysis or factor analysis, and the second is the problem of weighting the individual indicators.

Comparison of the extended Hellwig method with other techniques, such as the classical Hellwig, TOPSIS, and SAW methods, led us to certain main observations. All these techniques (TOPSIS, SAW, Hellwig) allow us to rank, group and compare countries from the perspective of sustainable development. Now, we pay attention to a few differences in the mentioned procedures. In Table 1, a comparison of those techniques is presented, which shows the similarities and differences among them as well as the added value of the extended Hellwig algorithm. First, comparing the Hellwig method to TOPSIS allows us to take into account only the pattern of development (an ideal vector), contrary to TOPSIS, where the pattern (ideal) and anti-pattern (anti-ideal) of development are considered. In particular, it is important in the context of such a situation when only the pattern is determined. It is noteworthy that the anti-pattern is very often artificially created. Second, the extended Hellwig procedure gives the opportunity to consider the individual pattern (based on national or EU target points) compared with the classical Hellwig method. Third, the algorithm of the Hellwig method shows the

Table 1 Comparison of the properties of the SAW, TOPSIS and Hellwig methods

Properties	Method			
	SAW	Classical TOPSIS	Classical Hellwig	Extended Hellwig
Pattern of development				
Based on max and min values	–	+	+	+
Based on EU target points	–	–	–	+
Based on national target points	–	–	–	+
Anti-pattern of development				
Based on max and min values	–	+	–	–

Key: +property is fulfilled; –property is not fulfilled

objects' distance from the pattern of development. Fourth, however, SAW (Churchman and Ackoff 1954) is very popular in measuring sustainable development, and it does not include the pattern of development or distances to them.

3.4 Sensitivity Analysis

We apply the proposed framework based on the extended Hellwig method to test its relevance and effectiveness in assessing the education component of the EU's agenda. For the sensitivity analysis of the results, we compare the classification of ranking results obtained by Hellwig method with those obtained by the Education Index.

The Education Index is a part of the well-known Human Development Index (HDI). The HDI is a summary measure of average achievement in key dimensions of human development: a long and healthy life, the ability to acquire knowledge and the achievement of a decent standard of living. Life expectancy at birth reflects the ability to lead a long and healthy life. The mean years of schooling and expected years of schooling reflect the ability to acquire knowledge. The gross national income per capita reflects the ability to achieve a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions (Human Development Report 2016, p.3). The Education Index (*EI*) is calculated using the following formula:

$$EI = \frac{MYSI + EYSI}{2}$$

The Mean Years of Schooling Index (*MYSI*) and Expected Years of Schooling Index (*EYSI*) are calculated in the following way:

$$MYSI = \frac{MYS}{15}, EYSI = \frac{EYS}{18}$$

Expected Years of Schooling (*EYS*) is the number of years of schooling that a child of school entrance age can expect to receive if prevailing patterns of age-specific enrolment rates persist throughout the child's life. Mean Years of Schooling (*MYS*) is the average number of years of education received by people aged 25 and older, converted from education attainment levels.

Next, we compare the results of grouping countries obtained by Hellwig method with the results of Ward's agglomeration method. Cluster analysis is a multivariate statistical technique that entails the division of a large group of objects into smaller and more homogeneous groups—clusters. For the purposes of this paper, one of the most commonly used methods, the agglomerative hierarchical clustering Ward method (Ward 1963), has been applied. Ward method is based on least squares criteria and minimizes the within-cluster sum of squares, thus maximizing the within-cluster homogeneity (for details, see Everitt et al. 2011; Romesburg 1984; among others).

Finally, the Mann–Whitney or Kruskal–Wallis tests can be used to identify indicators that are at significantly different levels in clusters. Those tests are rank-based nonparametric tests that can be used to determine every variable if there are statistically significant differences in averages for obtained classes (the Mann–Whitney test for two classes and the Kruskal–Wallis test for more than two classes).

4 Empirical Case Study Using the Extended Hellwig Procedure to Measure Sustainable Development in the Education Area

In this section, we describe an empirical case analysis to demonstrate the usability and effectiveness of the proposed framework for measuring sustainable development in the education area. The extended Hellwig method was applied in two cases, taking into consideration EU targets (*H2*) and national targets (*H3*). The classical Hellwig approach of the pattern of development with the modification of the normalization stage is denoted by *H1*.

4.1 Problem Description

A strategy for smart, sustainable and inclusive growth, known as the Europe 2020 strategy, was introduced by the European Commission (EC) in 2010 (EC 2010). The main aim of EC was to create a solid foundation for development based on three mutually reinforcing priorities:

- *Smart growth*: developing an economy based on knowledge and innovation,
- *Sustainable growth*: promoting a more resource efficient, greener and more competitive economy,
- *Inclusive growth*: fostering a high-employment economy delivering social and territorial cohesion (EC 2010, p.6). Two headline targets assumed in the Europe 2020 strategy correspond to both the education area and *smart growth*:
 - The share of early school leavers should be under 10%,
 - At least 40% of the younger generation should have a tertiary degree.

Reflecting different national circumstances, the two mentioned EU targets have also been transposed into national targets. Therefore, each EU country can check its own

progress towards each goal. The SDGs indicator set, which constituted the basis for assessing the level of implementation of the Europe 2020 strategy, as well as the EU targets identified in this strategy and national targets accepted by the member states, are collected in the Eurostat database.

Education¹, as a fundamental human right and an essence of intelligent development, also became one of the seventeen goals of *The 2030 Agenda for Sustainable Development* (known as *Agenda Post-2015*), which has given a new direction for achieving sustainable development (United Nations 2015). This goal, known as *Quality Education* (SDG 4), “seeks to ensure people have access to equitable and quality education through all stages of life, from early childhood education and care, through primary and secondary schooling, to technical vocational training and tertiary education” (Eurostat 2018). The indicators from the subset SDG 4 can be applied to assess the progress of EU countries towards SDG 4 on ensuring an inclusive and quality education for all. Thus, research on the education performance of European countries should be considered an important field of interest. The importance of this issue and the major impact of education on other areas connected with sustainable development have been explored by many authors in various contexts (Albareda-Tiana et al. 2018; De Wit et al. 2015).

4.2 Data Source

For the purpose of the study, the data made available by the Eurostat for the Sustainable Development Indicators in the area of education (SDG 4) were used. The SDG 4 indicator set consists of eight indicators that allow us to monitor progress on all levels and areas of education. The list of the indicators with EU targets used in the analysis is presented in Table 2. The comprehensive definitions of the indicators can be found in Eurostat 2018.

The Europe 2020 education targets are monitored with two headline indicators: *early leavers from education and training* (C1) and *tertiary educational attainment* (C2). As mentioned above, the main aim of this strategy was to reduce the share of C1 to less than 10% and increase the share of C2 to at least 40% by 2020. These two Europe 2020 education targets also feature EU benchmarks under the Strategic Framework for ET 2020. ET 2020 (EC 2018) takes into consideration other quantitative benchmarks that should be completed by 2020 and that are monitored with the following indicators (Eurostat 2018):

- *Participation in early childhood education* (C3)—at least 95% of children between the age of 4 and the age of starting compulsory primary education should participate in early childhood education.
- *Underachievement in reading, maths, and science* (C4–C6)—the share of low-achieving 15 year-olds in reading, mathematics and science should be less than 15%.
- *Employment rates of recent graduates* (C7)—the share of graduates (20–34 year-olds) having left education and training in the past 1–3 years who are employed and not in any further education and training should be at least 82%.
- *Adult participation in learning* (C8)—an average of at least 15% of adults should participate in learning.

¹ Education is defined as a “process of teaching, training and learning, to improve knowledge and develop skills” (Oxford Dictionary 2005, p.488). The United Nations made a declaration in its Agenda 21, where it is stated that “education is critical for promoting sustainable development and improving the capacity of the people to address environment and development issues” (United Nations 1992).

Table 2 List of SDG 4 indicators and EU targets

No	Indicator (characteristic of the indicator)	EU target
C1	<i>Early leavers from education and training</i> —% of population aged 18 to 24 with at most lower secondary education [sdg_04_10] (N)	Below 10%
C2	<i>Tertiary educational attainment</i> —% of population aged 30 to 34 [sdg_04_20] (P)	At least 40%
C3	Participation in early childhood education—% of the age group between 4 years of age and the starting age of compulsory education [sdg_04_30] (P)	At least 95%
C4	Underachievement in reading—% of 15-year-old students [sdg_04_40] (N)	Less than 15%
C5	Underachievement in maths—% of 15-year-old students [sdg_04_40] (N)	Less than 15%
C6	Underachievement in science—% of 15-year-old students [sdg_04_40] (N)	Less than 15%
C7	Employment rates of recent graduates—% of population aged 20 to 34 [sdg_04_50] (P)	At least 82%
C8	Adult participation in learning—% of population aged 25 to 64 [sdg_04_60] (P)	At least 15%

Source: Eurostat (EU-LFS 2016 for 1, 2, 5 and 6; UOE 2015 for 3) & OECD (PISA 2015 for 4).

NOTE: ISCED 0=early childhood education; 1=primary education; 2=lower secondary education; 3=upper secondary education; 4=post-secondary non-tertiary education; 5=short-cycle tertiary education; 6=bachelor's or equivalent level; 7=master's or equivalent level; 8=doctoral or equivalent level. P denotes a positive indicator, while N denotes a negative indicator

Monitoring SDG 4 in the EU context focuses on the issues of basic, tertiary and adult education. The set of all SDG 4 indicators allows us to look at educational attainment in general and at their impacts on the labour market. The indicators enable us to go through all stages of the education process beginning with early school leaving, followed by the typical educational pathway starting with early childhood education, through the acquisition of basic skills (reading, maths and science), leading to tertiary education and adult participation in learning (Eurostat 2017, 2018).

4.3 Analysis of the Results

In this part, the empirical results concerning the evaluation of the implementation level of the Europe 2020 strategy recommendations in the education area among the EU member states in 2015 are presented. The characterized multiple-criteria analysis methods were applied for comparative research on the education area of EU countries in 2015.

In the first step of the research using the SDG 4 indicator dataset, the composite indices of sustainability related to education based on Hellwig procedure were created. Those indices were constructed by considering the idea of *evaluation of education goals*, *realizing EU education targets*, and *realizing national education targets*. In this context, we used three different Hellwig models. They were suitable for evaluating the education domain from a three-point of view pattern. The first was the classical approach, which was based on the maximum and minimum values of SDG 4 indicators (by using the *H1* method). The second one was based on EU targets (*H2* method). The last one was created by means of national targets (*H3* method).

The education goals, defined in the Europe 2020 strategy, are measured by SDG 4 indicators (see Table 2). An initial set of diagnostic features has been verified by taking into consideration discriminative ability and informational potential. Although we found a strong correlation between *underachievement in reading, maths, and science*, we did not decide to eliminate any of these indicators. Finally, we include all indicators (Table 2) proposed by Eurostat by applying the idea of a comprehensive description of the education phenomenon by introducing an appropriate load of information into the analysis and sufficiently differentiating classified objects. Table 3 presents the final dataset for the 28 EU countries in 2015, consisting of eight indicators designated by symbols from *C1* to *C8*, together with national targets for *C1* and *C2*, and the main statistical parameters. Based on this table, we can make four observations. First, national targets for *early leavers from education and training* range from 4% for Croatia to 16% for Italy and from 26% for Italy to 66% for Luxembourg in 2015. Second, national targets for *early leavers from education and training* range from 4% for Croatia to 16% for Italy and from 26% for Italy to 66% for Luxembourg in 2015. Third, it shows that in 2015, eleven countries had already achieved their national targets for *C1* and thirteen countries had achieved it for *C2*. Finally, the EU countries are the most differentiable with respect to *adult participation in learning (C8)* and the least differentiable with respect to *participation in early childhood education (C3)*.

Table 4 summarizes the values of synthetic measures and rankings of the EU countries obtained by using the classical Hellwig method (*H1*) and the Hellwig procedures based on EU targets (*H2*) and national targets (*H3*), all with equal weights [see (Maggino and Ruviglioni 2009)].

Based on Table 4, we can make three main observations. First, taking into account the composite index for *H1*, we found that the country characterized by the highest level of achievement of the education goals in 2015 is Denmark, with a value of 0.757. The next

Table 3 The data for the 28 EU countries and national targets in 2015

Country	<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	National Target <i>C1</i>	National Target <i>C2</i>
Austria	7.3*	38.7	94.8	21.8	22.5	20.8	86.9*	14.4	9.5	38.0
Belgium	10.1	42.7*	98.3*	20.1	19.5	19.8	79.5	6.9	9.5	47.0
Bulgaria	13.4	32.1	89.2	42.1	41.5	37.9	74.6	2.0	11.0	36.0
Croatia	2.8*	30.8	73.8	32.0	19.9	24.6	62.9	3.1	4.0	35.0
Cyprus	5.2*	54.5*	89.6	42.6	35.6	42.1	68.9	7.5	10.0	46.0
Czech Republic	6.2*	30.1	88.0	21.7	22.0	20.7	82.2*	8.5	5.5	32.0
Denmark	7.8*	47.6*	98.5*	13.6*	15.0	15.9	81.7	31.3*	10.0	40.0
Estonia	12.2	45.3*	91.9	11.2*	10.6*	8.8*	80.4	12.4	9.5	40.0
Finland	9.2*	45.5*	83.6	13.6*	11.1*	11.5*	75.5	25.4*	8.0	42.0
France	9.2*	45.0*	100.0*	23.5	21.5	22.1	72.4	18.6*	9.5	50.0
Germany	10.1	32.3	97.4*	17.2	16.2	17.0	90.4*	8.1	10.0	42.0
Greece	7.9*	40.4*	79.6	35.8	27.3	32.7	45.2	3.3	10.0	32.0
Hungary	11.6	34.3	95.3*	28.0	27.5	26.0	80.4	7.1	10.0	34.0
Ireland	7.0*	51.9*	97.7*	15.0	10.2*	15.3	77.7	6.3	8.0	60.0
Italy	14.7	25.3	96.2*	23.3	21	23.2	48.5	7.3	16.0	26.0
Latvia	9.9*	41.3*	95.0*	21.4	17.7	17.2	78.8	5.7	10.0	34.0
Lithuania	5.5*	57.6*	90.8	25.4	25.1	24.7	82.1*	5.8	9.0	48.7
Luxembourg	9.3*	52.3*	96.6*	25.8	25.6	25.9	84.7*	18.0*	10.0	66.0
Malta	19.8	27.8	100.0*	29.1	35.6	32.5	95.1*	7.2	10.0	33.0
Netherlands	8.2*	46.3*	97.6*	16.7	18.1	18.5	88.2*	18.9*	8.0	40.0
Poland	5.3*	43.4*	90.1	17.2	14.4*	16.3	77.4	3.5	4.5	45.0
Portugal	13.7	31.9	93.6	23.8	17.2	17.4	72.2	9.7	10.0	40.0
Romania	19.1	25.6	87.6	39.9	38.7	38.5	68.1	1.3	11.3	26.7
Slovakia	6.9*	28.4	78.4	27.7	32.1	30.7	75.2	3.1	6.0	40.0
Slovenia	5.0*	43.4*	90.5	16.1	15.1	15.0	71.5	11.9	5.0	40.0
Spain	20.0	40.9*	97.7*	22.2	16.2	18.3	65.2	9.9	15.0	44.0
Sweden	7.0*	50.2*	95.0*	20.8	18.4	21.6	85.9*	29.4*	7.0	45.0
United Kingdom	10.8	47.9*	100.0*	21.9	17.9	17.4	85.7*	15.7*	10.0	40.0
<i>Mean value</i>	9.8	40.5	92.4	23.9	21.9	22.6	76.3	10.8		
<i>Standard deviation</i>	4.5	9.2	6.9	8.4	8.4	8.3	11.3	8.0		
<i>Variability coefficient</i>	45.5	22.8	7.4	35.2	38.5	36.5	14.8	74.5		

*The EU target is fulfilled

three positions were occupied by Sweden (0.703), the Netherlands (0.674) and the United Kingdom (0.610). These countries represented a very high level of education development. The regions characterized by the lowest level of achievement of the education goals are Bulgaria (0.054) and Romania (−0.021)—the last two positions in the ranking. Both of these countries are characterized by zero realized EU education targets. With the exception of these two countries and Portugal, the other EU countries have realized at least one goal.

Second, in terms of the rankings based on the Hellwig methods with EU targets (*H2*) and national targets for indicators (*H3*), we noticed a greater spread between the highest and the lowest values (1.012 for *H2* and 1.023 for *H3*). Both rankings again highlighted

Table 4 The values and rank ordering of EU countries obtained by the Hellwig measures and the Education Index

Country	$H1$	Range	$H2$	Range	$H3$	Range	Difference $H2-H1$	Difference $H3-H1$	Education Index	Range
Austria	0.539	11	0.801	7	0.802	7	0.262	0.263	0.847	15
Belgium	0.495	12	0.680	11	0.701	11	0.185	0.207	0.890	9
Bulgaria	0.054	27	0.097	27	0.097	27	0.043	0.043	0.811	21
Croatia	0.183	22	0.242	22	0.233	22	0.059	0.050	0.793	23
Cyprus	0.183	23	0.195	25	0.215	24	0.013	0.032	0.792	24
Czech Republic	0.403	17	0.649	14	0.563	18	0.246	0.159	0.891	8
Denmark	0.757	1	0.984	1	0.985	1	0.227	0.227	0.918	2
Estonia	0.573	7	0.838	5	0.854	4	0.265	0.280	0.869	12
Finland	0.605	5	0.737	8	0.747	9	0.132	0.143	0.904	5
France	0.553	9	0.736	9	0.759	8	0.183	0.206	0.840	16
Germany	0.476	15	0.644	15	0.672	14	0.169	0.196	0.940	1
Greece	0.127	26	0.149	26	0.169	26	0.021	0.042	0.833	17
Hungary	0.345	20	0.550	19	0.528	19	0.205	0.183	0.818	20
Ireland	0.585	6	0.686	10	0.712	10	0.101	0.127	0.910	4
Italy	0.214	21	0.295	21	0.281	21	0.081	0.066	0.791	25
Latvia	0.479	14	0.664	12	0.672	13	0.185	0.193	0.866	13
Lithuania	0.467	16	0.565	18	0.575	17	0.098	0.109	0.880	10
Luxembourg	0.548	10	0.649	13	0.693	12	0.102	0.145	0.788	26
Malta	0.160	25	0.205	24	0.195	25	0.045	0.035	0.810	22
Netherlands	0.674	3	0.915	2	0.918	2	0.242	0.244	0.903	6
Poland	0.492	13	0.593	17	0.605	16	0.101	0.113	0.853	14
Portugal	0.388	18	0.596	16	0.605	15	0.207	0.217	0.757	28
Romania	-0.021	28	-0.028	28	-0.039	28	-0.006	-0.017	0.767	27
Slovakia	0.178	24	0.209	23	0.229	23	0.031	0.051	0.831	18
Slovenia	0.564	8	0.806	6	0.811	6	0.242	0.247	0.879	11
Spain	0.348	19	0.448	20	0.466	20	0.099	0.118	0.819	19
Sweden	0.703	2	0.842	4	0.845	5	0.139	0.143	0.902	7
United Kingdom	0.610	4	0.864	3	0.867	3	0.254	0.257	0.911	3

Denmark (0.984 for H_2 ; 0.985 for H_3) as the region with a very high level of realization of education goals taking into consideration EU or national targets. The next two positions were taken by the Netherlands (0.915 for H_2 ; 0.918 for H_3) and the United Kingdom (0.864 for H_2 ; 0.867 for H_3). The last two places were allocated to the same regions as in the first ranking: Bulgaria (0.097 for H_2 and H_3) and Romania (-0.028 for H_2 and -0.039 for H_3).

Third, regarding the analysed positions of EU countries in the overall classification obtained by different Hellwig methods, it must be emphasized that some EU countries improved, while others worsened their positions in the ranking. Generally, the differences in values and rankings received by H_2 and H_3 are rather small. The largest difference in position in the rankings H_2 and H_3 concerns the Czech Republic, which dropped 4 positions (from 14th position for H_2 to 18th position for H_3). Fifteen countries did not change their positions. Comparing values of synthetic measures obtained by the classical Hellwig method (H_1) and by the modified methods taking into account EU or national targets (H_2 and H_3), it turned out that ten countries improved their positions, six countries did not change their positions in the rankings, but for all countries, the values of synthetic measures increased. Estonia (0.265 for H_2 and 0.280 for H_3), Austria (above 0.260) and the United Kingdom (above 0.250) improved their values the most, which translated into an increase in their positions in the rankings: Austria had the largest increase (4 positions), followed by Estonia (2 or 3 positions), and the United Kingdom (1 position). The smallest increase in the value of the indicator was recorded for Romania (-0.006 for H_2 and -0.017 for H_3), which caused it to have last position in all the rankings.

In summary, the results included in Table 4 demonstrate significant disparities in the level of implementation of the Europe 2020 strategy in the education area across EU countries.

We obtained the following classification with respect to H_1 :

Group 1 (*very high level of education development*): **Denmark, Sweden, the Netherlands;**

Group 2 (*medium-high level of education development*): the United Kingdom, **Finland, Ireland, Estonia, Slovenia, France, Luxembourg, Austria, Belgium, Poland, Latvia, Germany, Lithuania;**

Group 3 (*medium-low level of education development*): Czech Republic, Portugal, **Spain, Hungary, Italy;**

Group 4 (*very low level of education development*): **Croatia, Cyprus, Slovakia, Malta, Greece, Bulgaria, Romania.**

In the case of both methods (H_2 and H_3), independent of the used targets, we obtained the following classification:

Group 1 (*very high level of education development*): **Denmark, the Netherlands, the United Kingdom, Estonia, Sweden;**

Group 2 (*medium-high level of education development*): **Slovenia, Austria, Finland, France, Ireland, Belgium, Latvia, Luxembourg, Czech Republic, Germany, Portugal, Poland, Lithuania;**

Group 3 (*medium-low level of education development*): **Hungary, Spain, Italy;**

Group 4 (*very low level of education development*): **Croatia, Slovakia, Malta, Cyprus, Greece, Bulgaria, Romania.**

The countries from Group 1, which represented a very high level of education development, are characterized by a high (4 or 5) number of realized EU targets. The countries from groups 3 and 4, which were characterized by a medium-low and very low level of education development, reached at most two EU targets in 2015. The countries that

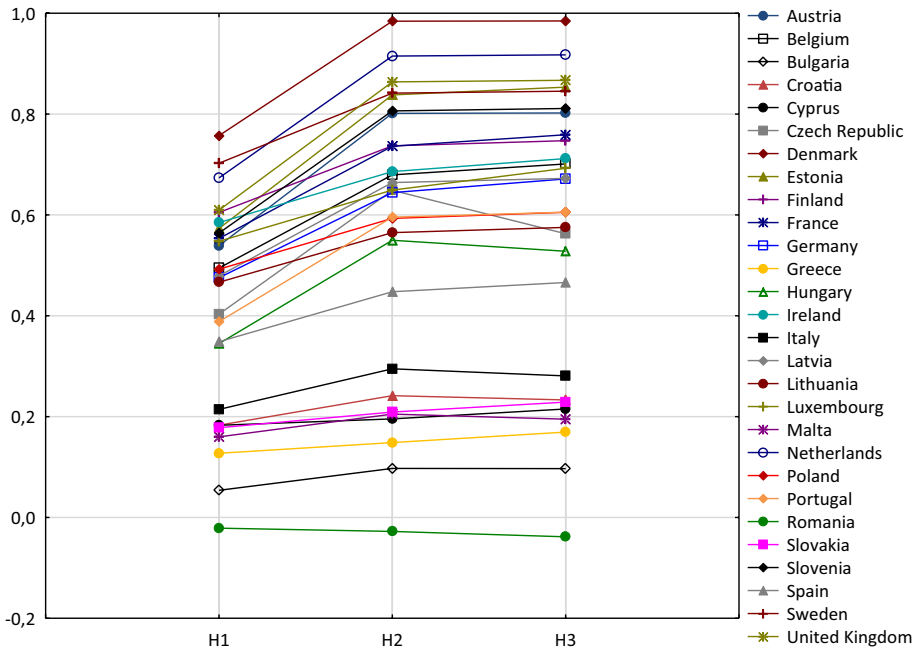


Fig. 2 Comparison of rankings obtained by means of the methods for H1-H3

appeared in the same group, regardless of the method used, are marked in bold. We can observe the following differences: The United Kingdom and Estonia are in Group 2 due to the *H1* method and in Group 1 due to the *H2* and *H3* methods used. The Czech Republic and Portugal appear in Group 3 due to the *H1* method and in Group 2 due to *H2* and *H3* methods. The fourth group includes the same countries in both classifications. The reshuffles among groups of both obtained classifications are not major.

Figure 2 confirms the diversity in education areas between EU countries. The range, in the case of the three applied methods, is as follows: 0.799 (for *H1*), 1.023 (for *H2*), and 1.012 (for *H3*). The Spearman’s correlation coefficients between the used methods are $r(H1, H2)=0.970$, $r(H1, H3)=0.971$, $r(H2, H3)=0.992$, and all of them are statistically significant ($p < 0.05$).

4.4 Comparative Analysis and Implications

The Hellwig procedures were proposed for ranking, classifying and comparing EU countries with respect to sustainable development in the education area in 2015. To illustrate that the proposed approach is effective, first, the results obtained using Hellwig method were compared with the results obtained using the Education Index. The values of the Education Index for EU countries in 2015 are presented in Table 3. Three observations from this table are as follows.

First, let us observe that the composite indices of the Education Index and Hellwig measure are based on different data. The indicators used in Hellwig procedure take into account all the processes of education, from early childhood to lower secondary, tertiary

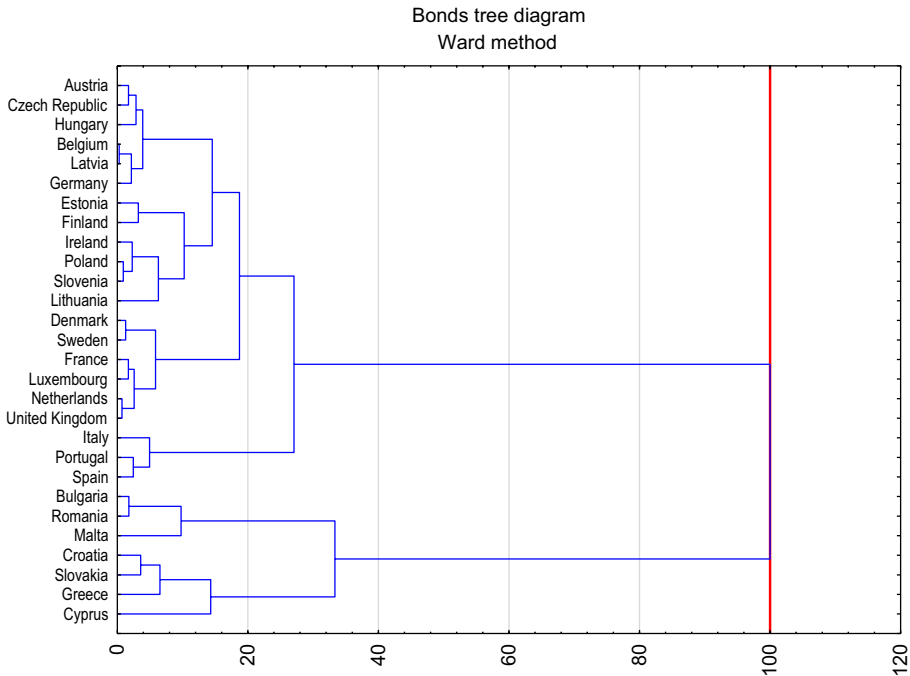


Fig. 3 Cluster analysis—bonds tree diagram Ward method for the EU-28 in 2015

and adult education. The main difference between these indices is that Hellwig measure takes into consideration the quality of knowledge (underachievement in maths, reading and science) as well as the employment rates of recent graduates, while the Education Index is based only on years of schooling. It seems that the proposed set of indicators $CI-C8$ (see Table 1) reflects the level of education from the perspective of sustainable development better than the given indicators.

Second, the Spearman's correlation coefficients among the rankings obtained by the used methods are $r(H1, EI)=0.726$, $r(H2, EI)=0.713$, $r(H3, EI)=0.685$, and all of them are statistically significant ($p < 0.05$). It is not surprising that the “quantitative” index does not correspond well with the “quality” of education represented by indices $H1$, $H2$, and $H3$ (Spearman's coefficient less than 0.730).

Third, the Education Index do not illustrate the variation among countries as Hellwig measure. The minimum value of the Education Index is 0.757 for Portugal, while the maximum value is 0.940 for Germany. To compare this result, for instance, with $H1$, we see that the minimum value is -0.021 for Romania and the maximum value is 0.757 for Denmark. For the other Hellwig indices, the differences between the minimum and maximum values are even greater.

In the paper (Roszkowska and Filipowicz-Chomko 2019), the extended TOPSIS procedure was proposed, which takes into account EU targets and/or national targets in building PIS and NIS. This algorithm allows compensatory and noncompensatory approaches in integrated sustainability assessment from the target point of view. This framework has been applied to measure sustainable development in the field of education using the same set of indicators, $CI-C8$. The Spearman's correlation coefficients between rankings obtained by the Hellwig and TOPSIS methods are greater than 0.980 (and are statistically significant

Table 5 Results of clustering

Clusters	Countries							
Cluster 1	Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Poland, Portugal, Slovenia, Spain, Sweden, United Kingdom							
Variables (average value)								
<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	
9.529	42.567	94.695	20.014	18.229	18.733	78.443	13.086	
Cluster 2	Bulgaria, Croatia, Cyprus, Greece, Malta, Romania, Slovakia							
Variables (average value)								
<i>C1</i>	<i>C2</i>	<i>C3</i>	<i>C4</i>	<i>C5</i>	<i>C6</i>	<i>C7</i>	<i>C8</i>	
10.729	34.229	85.457	35.600	32.957	34.143	70.000	3.929	

$p < 0.050$). This result confirms that both the Hellwig and TOPSIS methods can be applied to measure sustainable development by taking into account target points. We should remember that Hellwig procedure is based on the pattern of development, while TOPSIS is based on the concept of ideal and anti-ideal solutions.

In the last part of the study, cluster analysis was applied. To distinguish groups of similar (in terms of the values of the referred eight indicators) countries, Ward agglomeration method was used. The results of the cluster analysis are presented in the form of a tree diagram (see Fig. 3).

Based on the above analysis, two clusters have been identified. The results concerning the classified countries and the average values of individual variables in the distinct groups are included in Table 5.

The first group included 21 countries with a high or medium level of education development. The fourteen countries included in Cluster 1 are the countries of the so-called “old EU”, and the next seven are the countries that joined the EU in 2004. The second group, which includes so-called “new member states” in addition to Greece (old Union), form Cluster 2, with a low level of education development.

In the next step, the differences between clusters were verified. For these two clusters, the application of the test enables us to verify whether the values of the variables used in the research are significantly different for the considered classes. For these two clusters, the Mann–Whitney test confirmed that all variables (besides *C1*) differ significantly ($p < 0.05$). In conclusion, the results of this test indicate that the variables used in the multicriteria analysis were selected properly. Their values significantly differentiate the determined classes.

5 Conclusion

In this paper, we proposed a new methodological framework based on extended Hellwig procedures that allowed an integrated assessment of sustainability goals while considering target levels. The core of the classical Hellwig technique is a construction of synthetic indices that is designed as a distance from the abstract pattern of sustainable development (ideal solution). This abstract point is defined by means of maximum and minimum relations, and it is common for all objects (countries). In the extended Hellwig method, we

introduced the notion of individual patterns of development (taking into consideration EU targets and member states' individual targets). The proposed procedure enables us to create rankings of the countries as well as group the countries into four relatively homogeneous subsets (classes) with very high, medium-high, medium-low and very low levels of development.

This proposed framework was tested as part of the evaluation of the implementation of the Europe 2020 strategy, the EU's agenda for smart, sustainable and inclusive growth. The modified Hellwig procedure was used to analyse the implementation level of the Europe 2020 strategy recommendations and the spatial diversity of EU countries in the education area in 2015.

The results show the significant disparities in the implementation of the Europe 2020 strategy recommendations in the education area in EU countries in 2015. All the front-runners in the rankings, representing a high level of education development obtained by the Hellwig methods, are the "old" member states (Denmark, the Netherlands, the United Kingdom, and Sweden). The last positions in the rankings are taken by countries that joined the EU in 2007 (Bulgaria and Romania). On the other hand, among new member states, relatively high positions in the proposed rankings were taken by the Baltic countries (Estonia, Poland, Latvia, and Lithuania) and Central and Southern European countries such as Slovenia and the Czech Republic. Among the new member states, Estonia and Slovenia take the highest positions. Finally, among the old member states, only Greece is characterized by very low levels in the Hellwig measures.

We compared Hellwig approach with others using the Education Index, TOPSIS and Ward technique. We showed that the set of SDG 4 indicators is better for measuring sustainable development in the education area than the set of indicators used in the Education Index. We also showed that the results obtained by methodological approaches based on TOPSIS or Ward procedure are consistent.

The proposed modification of Hellwig method turned out to be an effective tool for assessing the implementation of the Europe 2020 strategy recommendations in the area of education in the EU countries. It is useful for assessing the relative performance towards a given target (national or/and EU target).

Our study also indicates new directions for research. In this work, all indicators were treated at an equal level, and equal weights were assumed. In further studies, it is worth considering the use of approaches based on different weights. These weights can be determined based on expert analysis or statistical methods. Furthermore, the usefulness of the proposed approach to monitor sustainable development in other areas based on the SDG set of indicators can be tested.

Acknowledgements Ewa Roszkowska was supported by the Polish Ministry of Science and Higher Education under subsidy for maintaining the research potential of the Faculty of Economy and Finance, University of Białystok, and Marzena Filipowicz-Chomko by the grant from Polish Ministry of Science and Higher Education (WZ/WI/1/2019).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albareda-Tiana, S., Vidal-Raméntol, S., & Fernández-Morilla, M. (2018). Implementing the sustainable development goals at University level. *International Journal of Sustainability in Higher Education*, 9(3), 473–497.
- Azapagic, A., & Perdan, S. (2005a). An integrated sustainability decision-support framework: Problem structuring, part I. *International Journal of Sustainable Development and World Ecology*, 12(2), 98–111.
- Azapagic, A., & Perdan, S. (2005b). An integrated sustainability decision-support framework: Methods and tools for problem analysis, part II. *International Journal of Sustainable Development and World Ecology*, 12(2), 112–131.
- Bandura, R. (2011). *Composite indicators and rankings: Inventory 2011*. Office of Development Studies, United Nations Development Programme (UNDP), New York: Technical report.
- Baster, N. (1972). *Measuring development: The role and adequacy of development indicators*. London: Frank Cass.
- Bilgin, M. H., Danis, H., Demir, E., Can, U. (2015). *Business Challenges in the Changing Economic Landscape - Vol. 1*. Proceedings of the 14th Eurasia Business and Economics Society Conference, Springer.
- Bolcárová, P., & Kološta, S. (2015). Assessment of sustainable development in the EU 27 using aggregated SD index. *Ecological Indicators*, 48, 699–705.
- Büyükoğkan, G., & Karabulut, Y. (2018). Sustainability performance evaluation: Literature review and future directions. *Journal of Environmental Management*, 217, 253–267.
- Churchman, C. W., & Ackoff, R. L. (1954). An approximate measure of value. *Journal of Operations Research Society of America*, 2(1), 172–187.
- Cinelli, M., Coles, S. R., & Kirwan, K. (2014). Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecological Indicators*, 46, 138–148.
- Colak, M. S., & Ege, A. (2013). An assessment of EU 2020 strategy: Too far to reach? *Social Indicators Research*, 110, 659–680.
- De Wit, H., Hunter, F., Howard, L., Egron-Polak, E. (2015). Internationalization of Higher Education. Study commissioned by Policy department B: Structural and Cohesion Policies. *Culture and Education*, European Parliament. [https://www.europarl.europa.eu/RegData/etudes/STUD/2015/540370/IPOL_STU\(2011\)540370_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2015/540370/IPOL_STU(2011)540370_EN.pdf). Accessed 29 January 2019.
- Diaz-Balteiro, L., González-Pachón, J., & Romero, C. (2017). Measuring systems sustainability with multi-criteria methods: A critical review. *European Journal of Operational Research*, 258(2), 607–616.
- Di Domizio, M. (2008). The competitive balance in the Italian football league: A taxonomic approach. *wpccommunity.it Department of Communication*, University of Teramo, 48.
- El Gibari, S., Gómez, T., & Ruiz, F. (2018). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics*, 89, 1–24.
- European Commission (2010). *Europe 2020 A strategy for smart, sustainable and inclusive growth*. Brussels: European Commission. <https://ec.europa.eu/eu2020/pdf/COMPLET%2520EN%2520BARROS%2520%2520%2520007%2520-%2520Europe%25202020%2520-%2520EN%2520version.pdf>. Accessed 29 January 2019.
- European Commission (2018). *Education and Training MONITOR 2018*, ec.europa.eu/education/monitor. Accessed 29 January 2019.
- Eurostat (2017). *Smarter, greener, more inclusive? INDICATORS TO SUPPORT THE EUROPE 2020 STRATEGY*, 2017 edition, <https://ec.europa.eu/eurostat/documents/3217494/8113874/KS-EZ-17-001-EN-N.pdf/c810af1c-0980-4a3b-bfdd-f6aa4d8a004e>. Accessed 29 January 2019.
- Eurostat (2018). *Europe 2020 indicators – education*, <https://ec.europa.eu/eurostat/statistics-explained/pdfscache/29305.pdf>. Accessed 29 January 2019.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis wiley series in probability and statistics*. UK: Wiley.
- Gana, X., Fernandezb, I. C., Guoc, J., Wilson, M., Zhaoe, Y., Zhou, B., et al. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81, 491–502.
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141, 61–94.
- Hellwig, Z. (1968). Application of the taxonomic method to the typological division of countries due to the level of their development and the structure of qualified personnel. *Statistical Review*, 4, 307–327.

- Hellwig, Z. (1972a). On the Optimal Choice of Predictors. In Z. Gostkowski (Ed.), *Towards a system of human capital resources indicators for less developed countries papers prepared for a UNESCO research project* (pp. 69–90). Wrocław: Ossolineum, Polish Academy of Sciences Press.
- Hellwig, Z. (1972b). Procedure of Evaluating High-Level Manpower Data and Typology of Countries by Means of the Taxonomic Method. In Z. Gostowski (Ed.), *Towards a system of human resources indicators for less developed countries, papers prepared for a UNESCO research project* (pp. 115–134). Wrocław: Ossolineum-The Polish Academy of Sciences.
- Huang, I. B., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of the total environment*, 409, 3578–3594.
- Human Development Report. (2016)
https://hdr.undp.org/sites/default/files/2016_human_development_report.pdf
- Hwang, C. L., & Yoon, K. (1981). *Multiple attribute decision-making: Methods and applications*. New York: Springer-Verlag.
- Janeiro, L., & Patel, M. K. (2015). Choosing sustainable technologies. Implications of the underlying sustainability paradigm in the decision-making process. *Journal of Cleaner Production*, 105, 438–446.
- Kou, G., Ergu, D., Lin, C. S., & Chen, Y. (2016). Pairwise comparison matrix in multiple criteria decision making. *Technol. Econ. Dev. Econ.*, 25, 738–765.
- Maggino, F., Ruvigliani E. (2009). *Obtaining weights: from objective to subjective approaches in view of more participative methods in the construction of composite indicators*. Seminar on New Techniques and Technologies for Statistics (NTTS) EUROSTAT, Brussels.
- Maggino, F. (2017). *Complexity in Society: From Indicators Construction to their Synthesis*. Cham: Springer.
- Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one. *Rivista Italiana di Economia Demografia e Statistica*, 67, 67–80.
- Mazziotta, M., & Pareto, A. (2017). Synthesis of Indicators: The Composite Indicators Approach. In F. Maggino (Ed.), *Complexity in society: From indicators construction to their synthesis* (pp. 159–191). New York: Springer.
- Munda, G. (2005a). Multi criteria decision analysis and sustainable development. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 953–986). New York: Springer.
- Munda, G. (2005b). Measuring sustainability: A multi-criterion framework. *Environment Development and Sustainability*, 7, 117–134.
- OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publishing.
- Dictionary, O. (2005). *Advanced learner's dictionary of current english* (7th ed.). New York: Oxford University Press.
- Pasimeni, P. (2012). Measuring Europe 2020: A new tool to assess the strategy. *International Journal of Innovation and Regional Development*, 4, 365–385.
- Pasimeni, P. (2013). The Europe 2020 index. *Social Indicators Research*, 110, 613–635.
- Pawlas, I. (2016). *Economic Picture of the Enlarged European Union in the Light of Taxonomic Research*. Proceedings of MAC-EMM 2016, 5th–6th August 2016 in Prague.
- Rappai, G. (2016). Europe en route to 2020: A new way of evaluating the overall fulfilment of the Europe 2020 strategic goals. *Social Indicators Research*, 129, 77–93.
- Reiff, M., Surmanová, K., Balcerzak, A. P., & Pietrzak, M. B. (2016). Multiple criteria analysis of european union agriculture. *Journal of International Studies*, 9(3), 62–74. <https://doi.org/10.14254/2071-8330.2016/9-3/5>.
- Rio Declaration on Environment and Development. (1992). https://www.unesco.org/education/pdf/RIO_E.PDF. Accessed 29 January 2019.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. Belmont: CA, Lifetime Learning Publications.
- Roszkowska, E. (2013). Rank ordering criteria weighting methods – a comparative overview. *Optimum Economic Studies*, 5(56), 14–33.
- Roszkowska, E., & Filipowicz-Chomko, M. (2019). Measuring sustainable development in the education area using multi-criteria methods: A case study. *Central European Journal of Operations Research*. <https://doi.org/10.1007/s10100-019-00641-0>.
- Rowley, H. V., Peters, G. M., Lundie, S., & Moore, S. J. (2012). Aggregating sustainability indicators: Beyond the weighted sum. *Journal of Environmental Management*, 111, 24–33.
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1), 83–98.

- Saaty, T. J., & Ergu, D. (2015). When is a decision-making method trustworthy? criteria for evaluating multi-criteria decision-making methods. *International Journal of Information Technology Decision Making*, *16*, 1171–1187.
- Saltelli, A., D’Hombres, B., Jesinghaus, J., Manca, A., Mascherini, M., Nardo, M., et al. (2011). Indicators for EU policies. Business as usual? *Social Indicators Research*, *102*, 197–207.
- Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological indicators*, *9*, 189–212.
- Talmaciu, A. M., & Cismas, L. M. (2016). National competitiveness through the Europe 2020 strategy and human development index in CEE countries. a panel data analysis. *Timis J Econ Bus*, *9*, 115–128.
- Ture, H., Dogan, S., & Kocak, D. (2018). Assessing Euro 2020 strategy using multi-criteria decision making methods: VIKOR and TOPSIS. *Social Indicators Research*. <https://doi.org/10.1007/s11205-018-1938-8>.
- UNESCO (1972a). *Synchronic and Diachronic Approaches in the UNESCO project on Human Resources Indicators*. Wroclaw Taxonomy and Bivariate Diachronic Analysis, SHC/WS/209 <https://unesdoc.unesco.org/ark:/48223/pf0000000801>. Accessed 29 January 2019.
- UNESCO (1972b). *Distance-based Analysis, Numerical Taxonomy and Classification of Countries according to selected areas of socio-economic development*, SHC/WS/237. <https://unesdoc.unesco.org/ark:/48223/pf0000001919>. Accessed 29 January 2019.
- United Nations (1992). Conference on Environment & Development, AGENDA 21, Rio de Janeiro Brazil. <https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf>. Accessed 29 January 2019.
- United Nations (2015). *Transforming our World: the 2030 agenda for sustainable development*, A/RES/70/1 <https://sustainabledevelopment.un.org/content/documents/21252030%2520Agenda%2520for%2520Sustainable%2520Development%2520web.pdf>. Accessed 29 January 2019.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association*, *58*(301), 236–244.
- Walheer, B. (2018). Decomposing the Europe 2020 Index. *Social Indicators Research*, *140*, 875–905.
- Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, *13*, 2263–3227.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.