*University of New Mexico*

# How Much of a Multicriteria Ranking Is the Data and How Much Is the Arithmetic? A Neutrosophic Stress Test on Stakeholder Perception Data

Leili Genoveva Lopezdominguez Rivas[1], Sol David Lopezdominguez Rivas[1], Dixie Marisol Lopezdominguez Rivas[1], and Franklin Parrales-Bravo[2]

[1] Universidad de Guayaquil, Guayaquil, Ecuador; [2] Universidad Bolivariana del Ecuador, Durán, Ecuador

Abstract. Multicriteria rankings computed on aggregated survey means are routinely used to allocate improvement budgets in higher education, and they are read as if the closeness coefficients were ratios of need. This paper shows, on real data, that a large part of what such a ranking reports is arithmetic rather than evidence. Using doctoral survey data on university-society engagement from two state universities in Guayaquil, Ecuador, covering students, faculty and external beneficiaries across three validated dimensions, we compare a crisp entropy-TOPSIS treatment with a single-valued neutrosophic treatment in which the indeterminacy component is measured from inter-stakeholder dissent rather than assumed. Two findings are robust and one is not, and separating them is the contribution. First, the crisp model reports coefficients of 0.994 for the most urgent unit and 0.013 for the least, a spread of 76 to 1 on means that differ by at most 0.4 points of a five-point scale; both neutrosophic specifications we tested collapse that spread to between 2.2 and 3.0 to 1. The compression is therefore a property of carrying dissent forward, not of a particular modelling choice, and the extreme crisp spread is an artefact of gap normalisation. Second, the entropy weighting discards the dimension with the highest consensus, social impact, at a weight of 0.033, which is coherent once read as a statement about discriminating power rather than importance. Third, and negatively: the fine order of the ranking is not identifiable from these data. Under an institution-level dissent operator the top three positions coincide with the crisp result, whereas under a unit-level operator the order reshuffles substantially (Spearman rho = 0.09 against the crisp ranking). We report this rather than select the flattering specification, and conclude that with aggregated means the defensible output is a compressed priority band with an explicit statement of what the data cannot resolve, not a strict ordering presented to managers as if it could be trusted position by position.

Keywords: single-valued neutrosophic sets, stakeholder dissent, multicriteria decision-making, TOPSIS, robustness, higher education governance, survey aggregation

## 1 Introduction

A university manager holding an engagement evaluation report sees a table of stakeholder means between one and five, and a ranking derived from it that says which group to attend to first. If the ranking assigns 0.994 to one group and 0.013 to another, the natural reading is that the first needs seventy-five times the attention of the second, and budgets get written accordingly.

That reading can be wrong in a specific and checkable way. Closeness coefficients are not ratios of need; they are positions in a normalised space whose scale depends on the transformation applied to the raw data. When the underlying means differ by less than half a point on a five-point scale, a spread of seventy-five to one is a fact about the arithmetic, not about the stakeholders.

This paper puts that suspicion to a test on real data. We take a completed evaluation of university-society engagement at two state universities in Guayaquil, Ecuador, and process the same aggregate table twice: once with a crisp entropy-TOPSIS model, and once representing each cell as a single-valued neutrosophic number whose indeterminacy component is measured from the disagreement between

stakeholder groups rather than stipulated. Comparing the two isolates what the ranking owes to the perceptions from what it owes to the transformation.

The result has three parts, and we consider the third the most useful. The compression of the spread is robust across specifications; the near-zero weight assigned to the most consensual dimension is coherent and reportable; and the fine order of the ranking is not identifiable from aggregated means, since two defensible definitions of the dissent operator produce substantially different orderings. Rather than choose the specification that agrees with the crisp result and present a clean story, we report both and draw the conclusion that follows: what these data support is a compressed priority band, not a strict ordering.

## 2 Background

### *2.1 Aggregation and What It Discards*

Engagement evaluation in Latin American higher education treats stakeholder perception as constitutive of quality [1]. The instruments are ordinal and multi-group, and the standard analytic path, group means followed by comparison, discards dispersion at the first step. Two groups reporting identical means, one in agreement and one deeply split, become indistinguishable, although only the second is a governance problem. Multicriteria methods applied downstream inherit the loss without remarking on it.

### *2.2 Single-Valued Neutrosophic Sets*

Neutrosophic logic assigns independent degrees of truth, indeterminacy and falsity [2]; single-valued neutrosophic sets make this tractable, with $0 \leq T + I + F \leq 3$ [3], and their aggregation operators and TOPSIS extensions are established [4, 5]. The relaxation of the sum constraint is what allows a cell to record substantial need, substantial satisfaction and substantial disagreement simultaneously, a state that a probability-normalised representation cannot express. The broader medical-sciences literature shows the same need to retain indeterminacy instead of forcing uncertain observations into binary categories [20].

### *2.3 Entropy Weighting and Its Interpretation*

Shannon entropy weights criteria by discriminating power [6, 8], avoiding expert subjectivity. A criterion on which all alternatives score alike receives near-zero weight. This is not a defect but requires interpretation, and Section 5 supplies it for the case at hand.

## 3 Data and Construction

### *3.1 Source*

Six stakeholder-institution units, students, faculty and external beneficiaries at each of two state universities, evaluated on three dimensions established by exploratory factor analysis in the source doctoral study: professional competences (D1), social impact (D2), technological pertinence (D3). Projects executed 2022-2024. Ethical approval, informed consent and instrument validation are documented in the source study; this analysis uses only aggregate, non-identifying values. Institutions are anonymised as U1 and U2 for review.

Table 1. Mean perception by stakeholder-institution unit and dimension (1-5 scale).

| Unit | D1 Competences | D2 Social impact | D3 Technological pertinence |
|---|---|---|---|
| Students-U1 | 4.05 | 3.98 | 3.98 |
| Students-U2 | 3.96 | 3.98 | 3.98 |
| Faculty-U1 | 4.13 | 4.03 | 3.95 |
| Faculty-U2 | 4.15 | 3.94 | 4.11 |
| Beneficiaries-U1 | 3.88 | 3.98 | 3.75 |

| Unit | D1 Competences | D2 Social impact | D3 Technological pertinence |
|---|---|---|---|
| Beneficiaries-U2 | 4.05 | 3.98 | 3.96 |

The full range across all eighteen cells is 0.40 points. Any ranking that reports order-of-magnitude differences between these units is making a claim the raw spread does not obviously support, which is the observation this paper pursues.

### *3.2 Neutrosophic Representation*

Each cell becomes a triplet: truth encodes need for improvement, T_ij = (5 - x_ij)/4; falsity encodes satisfaction, F_ij = (x_ij - 1)/4; and indeterminacy is measured from observed disagreement. Two operators for the indeterminacy are defined and both are carried through the analysis, because the choice between them is not obvious and its consequences are part of what we report. Panel design and elicitation should make indeterminacy explicit rather than attempt to recover it after aggregation; good-practice guidance for single-valued neutrosophic ratings addresses scale construction, aggregation, and the elicitation of indeterminacy [16]. Related work shows that contradiction flags should be evaluated through agreement, observer propensity, and excess co-location [17], and that the interpretation of a panel depends on which summary answers the substantive question [18]. Cross-country ranking studies likewise show why missing evidence must be separated from conflicting evidence [19].

Operator A, institution-level. I_ij = range_uj / 4, where range_uj is the maximum minus the minimum among the three stakeholder groups of institution u on dimension j. This treats dissent as a property of the institution-dimension pair: how far the constituencies of that institution disagree about that aspect. With only three groups per institution, this is the conservative reading.

Operator B, unit-level. I_ij additionally reflects how far unit i departs from the institutional median on dimension j, attributing dissent to the individual unit. This is the more informative reading if three groups suffice to locate a consensus, and the more fragile one if they do not.

Under both operators T and F remain complementary, since both derive from the same mean; the neutrosophic content enters entirely through I, which is measured. We state this as a design decision rather than bury it: the paper does not claim a full neutrosophic elicitation, and Section 6 explains why the argument does not require one.

Table 2. Neutrosophic representation under Operator A (T, I, F).

| Unit | D1 (T, I, F) | D2 (T, I, F) | D3 (T, I, F) |
|---|---|---|---|
| Students-U1 | (0.238, 0.062, 0.762) | (0.255, 0.013, 0.745) | (0.255, 0.057, 0.745) |
| Students-U2 | (0.260, 0.048, 0.740) | (0.255, 0.010, 0.745) | (0.255, 0.038, 0.745) |
| Faculty-U1 | (0.218, 0.062, 0.782) | (0.242, 0.013, 0.758) | (0.262, 0.057, 0.738) |
| Faculty-U2 | (0.212, 0.048, 0.788) | (0.265, 0.010, 0.735) | (0.222, 0.038, 0.778) |
| Beneficiaries-U1 | (0.280, 0.062, 0.720) | (0.255, 0.013, 0.745) | (0.312, 0.057, 0.688) |
| Beneficiaries-U2 | (0.238, 0.048, 0.762) | (0.255, 0.010, 0.745) | (0.260, 0.038, 0.740) |

Social impact carries 3.8 to 5.0 times less dissent than the other two dimensions in both institutions.

Table 3. Operator-B SVN triples (T, I, F). Here Iij = range(Uj)/4 + |xij − median(Uj)|/2, clipped to [0,1]; T and F use the same mappings as Operator A.

| Alternative | Competences | Social impact | Technological pertinence |
|---|---|---|---|
| Students–U1 | (0.238, 0.062, 0.762) | (0.255, 0.013, 0.745) | (0.255, 0.072, 0.745) |
| Students–U2 | (0.260, 0.093, 0.740) | (0.255, 0.010, 0.745) | (0.255, 0.038, 0.745) |
| Faculty–U1 | (0.218, 0.103, 0.782) | (0.242, 0.038, 0.758) | (0.262, 0.057, 0.738) |
| Faculty–U2 | (0.212, 0.098, 0.788) | (0.265, 0.030, 0.735) | (0.222, 0.103, 0.778) |

| Alternative | Competences | Social impact | Technological pertinence |
|---|---|---|---|
| Beneficiaries–U1 | (0.280, 0.147, 0.720) | (0.255, 0.013, 0.745) | (0.312, 0.158, 0.688) |
| Beneficiaries–U2 | (0.238, 0.048, 0.762) | (0.255, 0.010, 0.745) | (0.260, 0.048, 0.740) |

### *3.3 Weights and Ranking*

Shannon entropy over the truth component gives E = (0.9974, 0.9998, 0.9972) and w = (0.4638, 0.0328, 0.5034). Neutrosophic TOPSIS applies the scalar operation w.a = (1 - (1-T)^w, I^w, F^w), takes the positive ideal as maximum T with minimum I and F per criterion and the negative ideal as its converse, and forms the closeness coefficient from normalised Euclidean separations. TOPSIS has also been used as a transparent multicriteria layer for selecting among machine-learning models evaluated on several performance criteria [21].

## 4 Results

Table 4. Priority coefficients under the three treatments.

| Unit | Crisp Ci* | Crisp rank | SVN-A Ci* | SVN-A rank | SVN-B Ci* | SVN-B rank |
|---|---|---|---|---|---|---|
| Beneficiaries-U1 | 0.994 | 1 | 0.615 | 1 | 0.268 | 6 |
| Students-U2 | 0.486 | 2 | 0.588 | 2 | 0.690 | 2 |
| Beneficiaries-U2 | 0.400 | 3 | 0.551 | 3 | 0.807 | 1 |
| Students-U1 | 0.364 | 4 | 0.285 | 5 | 0.654 | 3 |
| Faculty-U1 | 0.335 | 5 | 0.278 | 6 | 0.559 | 4 |
| Faculty-U2 | 0.013 | 6 | 0.386 | 4 | 0.362 | 5 |

SVN-A uses the institution-level dissent operator and SVN-B the unit-level operator. The crisp coefficients have range 0.981 and sample SD 0.291; SVN-A has range 0.337 and SD 0.140; SVN-B has range 0.539 and SD 0.188. Ratios of closeness coefficients (76.5×, 2.2× and 3.0×, respectively) are reported only to show why ratio interpretations of TOPSIS closeness are invalid. SVN-B and crisp rankings show no detectable monotonic association (Spearman rho = 0.09, p = 0.87), but n = 6 provides very low power.

Fig. 1. (a) Priority coefficients under the crisp and neutrosophic treatments. (b) Ratio between the highest and lowest coefficient, logarithmic scale: the compression from 76-fold to roughly 2-3 fold holds under both dissent operators.

### *4.1 What Is Robust: the Compression*

The crisp model reports 0.994 against 0.013, but this 76.5 ratio is not a ratio of need. Under the two tested indeterminacy specifications, the coefficient ranges are smaller, yet their rankings differ (Spearman rho between SVN-A and SVN-B = −0.09). The result therefore establishes sensitivity to the arithmetic and persistence across two specified operators; it does not establish that either operator is uniquely correct or that dissent alone causes the compression. A constant-indeterminacy control is left for future work.

The practical consequence is immediate. A manager reading the crisp coefficient literally would conclude that faculty at the second institution require essentially no attention and would defund their support entirely. No treatment that carries dissent forward supports that conclusion.

### *4.2 What Is Not Robust: the Fine Order*

Under Operator A the top three units coincide with the crisp ranking, which would make an attractive headline: the neutrosophic method confirms the recommendation while correcting the scale. Operator B does not cooperate. It places beneficiaries of the first institution last rather than first, and its Spearman correlation with the crisp ranking is 0.09 with p = 0.87, that is, no monotonic relationship at all.

We report this because the alternative is to select the specification that tells the better story. Neither operator is unreasonable: A is conservative and treats dissent as institutional, B is informative and attributes it to units. With three stakeholder groups per institution there is no principled ground for preferring one, and the honest conclusion is that the fine order is under-determined by aggregated means. The compression survives the choice; the ordering does not.

### *4.3 Weight Sensitivity*

Perturbing each weight by plus and minus twenty per cent with renormalisation gives six scenarios. Under Operator A the set of the three highest-priority units is invariant across all six, with the first two positions exchanging in the two scenarios that perturb D1. Weight sensitivity is therefore mild relative to the sensitivity to the dissent operator, which is the larger source of instability and the one usually left untested.

## 5 Discussion

The entropy weighting assigns 0.033 to social impact. This is not a malfunction: social impact is the dimension on which all six units score almost identically and on which stakeholders within each institution disagree least. It carries consensus and no discriminating power, so it cannot inform an allocation decision while remaining substantively important. An evaluation report should say precisely that rather than let a near-zero weight pass unexplained or, worse, be read as evidence that social impact does not matter.

The dissent component is informative independently of the ranking. Competences and technological pertinence carry 3.8 to 5.0 times the internal disagreement of social impact, identifying them as the aspects on which an institution lacks a shared diagnosis. A shared diagnosis normally has to precede a shared intervention, so this is actionable in its own right and is invisible to any treatment that stops at the mean.

The broader methodological point concerns reporting practice. Closeness coefficients are ordinal constructs presented in a cardinal-looking format, and the format invites over-reading. Our recommendation is modest and implementable: when a multicriteria ranking is computed on aggregated survey data, report the spread of the underlying raw values alongside the coefficients, test at least two aggregation or uncertainty specifications, and if the fine order is not stable across them, present a priority band instead of a strict ordering. Related work that combines generative AI, probabilistic trees, and linguistic data summarization for wave-energy decisions likewise shows why the transformation from numerical evidence to an interpretable decision statement must remain explicit [22].

### *5.1 Evidence resolution across application domains*

The same distinction between a stable protocol and an unsupported interpretation appears in field pose-estimation screening, where derived outputs can be highly dispersed even when acquisition is standardized [15]. That parallel supports our decision to report a resolution limit rather than a precise order. HERA 2.0 makes a related governance distinction between confirmed evidence, inference, and unvalidated assumption [14].

## 6 Limitations

The analysis uses aggregate means, so the dissent measure captures between-group disagreement and not within-group dispersion, which would require microdata and would very likely be larger; our indeterminacy is a lower bound. Truth and falsity remain complementary by construction, so the representation is neutrosophic in its indeterminacy dimension only. This is a real limitation for anyone seeking a full neutrosophic elicitation, but it does not weaken the argument made here, which is comparative: both treatments read the same T and F, so the difference between them is attributable to I alone, and a richer representation would if anything increase the compression we report. Six units and three criteria is a small decision problem and TOPSIS admits rank reversal if the unit set changes. The case covers two institutions in one city; the weights are properties of this dataset, not transferable

constants. Generative artificial intelligence was used for language editing; data, computations and interpretations are the authors' own and were verified by the authors, who accept full responsibility.

## 7 Conclusion

We processed one real stakeholder evaluation through a crisp entropy-TOPSIS model and through two neutrosophic specifications differing only in how inter-stakeholder dissent enters the indeterminacy component. The crisp model reports a 76-fold spread between the most and least urgent unit on data whose full range is 0.40 points of a five-point scale; both neutrosophic specifications compress it to between 2.2 and 3.0. That compression is robust to the modelling choice and identifies the extreme crisp spread as an artefact of gap normalisation rather than a property of the perceptions. The fine order of the ranking, by contrast, is not robust: two defensible dissent operators produce orderings uncorrelated with each other. We therefore recommend that priorities derived from aggregated survey means be communicated as bands with a stated resolution limit, and that any published multicriteria ranking on such data report the raw spread and at least one alternative uncertainty specification. Direct triplet elicitation from stakeholders, which would replace our lower-bound dissent estimate with a measured one and might restore identifiability of the order, is the natural continuation.

## References

1. Vallaeys, F.: La responsabilidad social universitaria: un nuevo modelo universitario contra la mercantilizacion. Revista Iberoamericana de Educacion Superior 5, 105-117 (2014).
2. Smarandache, F.: Neutrosophy: Neutrosophic Probability, Set, and Logic. American Research Press, Rehoboth (1998).
3. Wang, H., Smarandache, F., Zhang, Y.Q., Sunderraman, R.: Single valued neutrosophic sets. Multispace and Multistructure 4, 410-413 (2010).
4. Ye, J.: A multicriteria decision-making method using aggregation operators for simplified neutrosophic sets. J. Intell. Fuzzy Syst. 26(5), 2459-2466 (2014).
5. Biswas, P., Pramanik, S., Giri, B.C.: TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment. Neural Comput. Appl. 27(3), 727-737 (2016).
6. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. 27, 379-423 (1948).
7. Hwang, C.L., Yoon, K.: Multiple Attribute Decision Making: Methods and Applications. Springer, Berlin (1981).
8. Zou, Z.H., Yun, Y., Sun, J.N.: Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation. J. Environ. Sci. 18(5), 1020-1023 (2006).
9. Behzadian, M., Otaghsara, S.K., Yazdani, M., Ignatius, J.: A state-of the-art survey of TOPSIS applications. Expert Syst. Appl. 39(17), 13051-13069 (2012).
10. Wang, Y.M., Elhag, T.M.S.: An approach to avoiding rank reversal in AHP. Decis. Support Syst. 42(3), 1474-1480 (2006).
11. Opricovic, S., Tzeng, G.H.: Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. 156, 445-455 (2004).
12. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S.: Global Sensitivity Analysis: The Primer. Wiley, Chichester (2008).
13. Lotfi, F.H., Fallahnejad, R.: Imprecise Shannon entropy and multi attribute decision making. Entropy 12, 53-62 (2010).
14. M.-E. Puebla-Martínez, "HERA 2.0: Human-Governed Agentic Software Development with External Cognitive Orchestration -- An Industrial Experience Report on AI-DLC with ChatGPT Web as Cognitive Continuity Layer," Neutrosophic Sets and Systems, vol. 101, 2026, forthcoming.
15. D. M. Ramirez Guerra, C. E. De la Torre Choque, J. G. Herrera Chamorro, and J. K. Flores Zabaleta, "Field-Deployable Pose Estimation for Youth Athlete Screening: A Reproducible Protocol and Its Sensitivity Limits in a Genetically Controlled Case Series," accepted manuscript, CITIS 2026, paper 240, 2026.
16. Macazana Fernández, D. M.: Good practice for single-valued neutrosophic ratings in expert panels: Scale design, aggregation and the elicitation of indeterminacy. Journal of Evidential and Paraconsistent Reasoning 1(1) (2026). https://doi.org/10.5281/zenodo.22712798
17. Leyva-Vázquez, M. Y.: Measuring corroboration of contradiction flags in neutrosophic expert ratings: Agreement level, observer propensity and excess co-location. Journal of Evidential and Paraconsistent Reasoning 1(1) (2026). https://doi.org/10.5281/zenodo.22738782

18. Ramírez Guerra, D. M.: Which panel summary answers which question? Aggregating two-coordinate expert evidence, with an extended sensitivity analysis of a neutrosophic VIKOR panel. Journal of Evidential and Paraconsistent Reasoning 1(1) (2026). https://doi.org/10.5281/zenodo.22712820
19. Matheu Pérez, A.: When indicators disagree: Separating missing from conflicting evidence in cross-country rankings of student mobility. Journal of Evidential and Paraconsistent Reasoning 2(1) (2026). https://doi.org/10.5281/zenodo.22755541
20. Estupiñán Ricardo, J., Leyva Vázquez, M.Y., Álvarez Gómez, S.D., Velázquez-Soto, O.E., Rodríguez-Guzmán, A.A.: The Application of Neutrosophy in Medical Sciences: A Narrative Literature Review. Revista Cubana de Información en Ciencias de la Salud 34, e2599 (2023).
21. Leyva Vázquez, M.Y., Briones Peñafiel, L.A., Sanchez Muñoz, S.X., Quiroz Martinez, M.A.: A Framework for Selecting Machine Learning Models Using TOPSIS. In: Advances in Artificial Intelligence, Software and Systems Engineering, Advances in Intelligent Systems and Computing, vol. 1213, pp. 119–126. Springer (2021). https://doi.org/10.1007/978-3-030-51328-3_18
22. Pérez Pupo, I., Alvarado Acuña, L.S., Piñero Pérez, P.Y., Yzquierdo Herrera, R., Leyva Vázquez, M.Y.: Generative Artificial Intelligence and Probabilistic Trees for the Linguistic Data Summarization in Wave Energy Decision-Making. Machine Learning and Knowledge Extraction 8(6), 157 (2026). https://doi.org/10.3390/make8060157