
RESEARCH ARTICLE

TWO PROVINCES OF THIRTY-EIGHT: WHEN A CORRELATION IS AN ARTEFACT OF ITS OUTLIERS

Yahya Syahrizal, S.Pd.^{1*}

¹ Semarang State University, Semarang, Indonesia

* Corresponding author: yahyasyahrizal15@gmail.com

Received 31 January 2026 | Revised 24 February 2026 | Accepted 21 March 2026

ABSTRACT

A correlation of 0.4893 between provincial primary and tertiary-age school participation in Indonesia falls to 0.0178 when two of the thirty-eight provinces are removed. This article uses that case to show how a cheap and rarely reported diagnostic identifies such failures. Using Central Statistics Agency figures for 2025 and 2023 covering all thirty-eight provinces, it computes both Pearson and Spearman correlations for four relationships in Indonesian education data. Three behave normally: participation at ages sixteen to eighteen against ages nineteen to twenty-three gives 0.7337 and 0.6615, primary against upper secondary completion gives 0.6725 and 0.6067, and the two dimensions of the national literacy index give 0.9462 and 0.8807. The fourth does not: primary against tertiary-age participation gives 0.4893 against a Spearman value of 0.0734, a gap of 0.4159. Deleting single observations confirms the diagnosis, with Papua Pegunungan alone moving the coefficient by 0.1489 and Papua Tengah by 0.1003; removing both leaves 0.0178. The article argues that a Pearson-Spearman gap of this size is a sufficient signal to investigate, that the appropriate response is to report both coefficients and the influential observations rather than to delete them, and that the influential provinces are usually the substantively important ones.

Keywords: *correlation; outliers; Spearman; educational statistics; measurement; Indonesia*

1. INTRODUCTION

Correlations between provincial education indicators are reported routinely and are used to support claims about how the parts of a system relate to one another. A positive correlation between participation at one stage and at another is read as evidence that provinces which do well early do well later.

The Pearson correlation coefficient is sensitive to observations far from the mass of the data, and where a small number of administrative units lie far below the rest — as they often do in a country with a wide development gradient — the coefficient can be produced almost entirely by them. The relationship it appears to describe may not hold among the remaining units at all.

This article documents an instance and uses it to demonstrate a diagnostic. Across the thirty-eight Indonesian provinces the Pearson correlation between participation at ages seven to twelve and at ages nineteen to twenty-three is 0.4893. The Spearman rank correlation over the same pair is 0.0734. Removing Papua Tengah and Papua Pegunungan, two of the thirty-eight, leaves a Pearson coefficient of 0.0178.

The diagnostic is the divergence between the two coefficients. Where a monotone relationship holds throughout the data, Pearson and Spearman are usually close; where a small number of extreme observations create the linear association, Pearson exceeds Spearman by a wide margin. The article computes both for four relationships in the same body of data and finds three of them well behaved and one not.

The research gap is that Indonesian provincial education correlations are published without either the rank coefficient or an influence diagnostic alongside them, so a reader cannot tell which case they are looking at. This article shows what the diagnostic costs, which is nothing, and what it reveals.

2. LITERATURE REVIEW

2.1 Measures that do not measure what they name

A first strand documents cases in which an educational statistic fails to represent its stated concept.

Fazlul, Koedel and Parsons (2023) show that enrolment in free and reduced-price school meals, used across the United States as a proxy for student poverty, does not measure student poverty, and trace the policy consequences. Viano, Baker, Ford and Johnston-Guerrero (2024) apply latent class analysis to racial terminology in education research and find the classifications used to be inconsistent across a single literature. Ford, Iantosca and Cate (2025) document the variability with which university students identify their own ethnoracial category.

Bokhove, Jerrim and Sims (2024) test whether schools employing an inspector receive better inspection grades, a design that treats an official measure as itself requiring validation. Goldhaber, Grout, Wolff and Martinkova (2021) assess the dimensionality and reliability of professional reference ratings and find both weaker than their use assumes.

The common lesson is that a published educational number carries a name and a definition, and that the relationship between them is an empirical question. The present article extends this to a derived statistic rather than a primary one.

2.2 Dimensionality and what a coefficient summarises

A second strand concerns the structure behind a summary number.

Min and Aryadoust (2021) review item response theory in language assessment and draw out the implications for whether language ability is one dimension or several. Stals, Isac and Claes (2024) evaluate multi-dimensionality and cross-national comparability in a measure of political trust among European youth, and find comparability to fail in places. Doz, Bone and Felda (2026) propose a multidimensional mathematics grade using fuzzy logic, and Coates (2026) fits a diagnostic classification model to an international mathematics assessment.

O'Neill, Lauridsen, Ostengaard and Qvortrup (2023) review validity evidence for a widely used teaching and learning questionnaire and find the evidence uneven. White and Maher (2024) ask what a standardised observation system adds to a summative teacher evaluation, which is the question of whether a second measure carries information beyond the first.

These studies establish that structure matters and is testable. A correlation coefficient is a summary of a joint distribution, and the diagnostic proposed here is a minimal test of whether the summary is representative.

2.3 Educational inequality across administrative units

A third strand supplies the substantive context in which such correlations are computed.

Wagner, Nair, Thapa and Kumar (2025) propose a Gini-based learning inequality index applicable across contexts, and Adeleke and McSharry (2022) show that a national education index is well predicted by a small number of indicators. Kelchen, Ortagus, Rosinger, Baker and Lingo (2023) relate state funding strategies to access and success across administrative units. Armand, Herman and Honore (2024) measure inequality of educational opportunity among primary school children in a single country.

Dearing, Zachrisson, Bustamante and Vandell (2026) examine the developmental timing of educational opportunity, and Sanfo and Malgoubri (2023) whether instructional quality narrows socioeconomic gaps. Both would ordinarily be reported with correlations of the kind examined here.

Ahmed (2023) argues that the field has been asking the wrong questions, and one small version of that argument is that a coefficient is often reported where a scatter would have been more honest.

2.4 Outliers as substance rather than nuisance

A fourth strand concerns what to do with observations that dominate a statistic.

The instinct to remove an influential observation is generally the wrong one in administrative data, because the influential unit is usually the one the analysis exists to understand. Prentice (2022) studies practices with refugee pupils at two English schools, and Jover, Manzano, Llos and Termes (2025) the experiences of families in a segregation-reduction programme in Barcelona; in both, the population at the tail is the subject.

Luschei and colleagues (2026) compare student well-being across Colombia, India and Malawi and treat the between-country differences as the finding rather than as noise. Serrano-Ardila, Baena-Navarro and Carriazo-Regino (2026) address digital integration specifically in rural education, again a setting defined by its distance from the mean.

Engel and Burch (2021) set out methodological considerations for policy sociology in a global era, among them the difficulty of comparing units that are not commensurable. Jiang, Ko-Wong and Valdovinos Gutierrez (2025) assess the feasibility of automated analysis in equity-focused research, where the units of interest are by construction unusual.

2.5 Stating what a measure is for

A fifth strand concerns the discipline of specifying, in advance, what a number is supposed to establish.

Gates, Teasdale, Shim and Hubacz (2024) argue for explicit specification of evaluative criteria in education, on the ground that an unstated criterion cannot be examined. The point applies directly to a correlation coefficient: a reader told that two provincial indicators correlate at 0.4893 is not told what claim the number is meant to support, and the two candidate claims — that provinces performing well early perform well later, and that two provinces perform badly at every stage — are answered differently by the same figure.

Yan and Pastore (2022) develop and validate a scale for teacher formative assessment literacy, an exercise in which the validation is the contribution rather than a preliminary to it. O'Neill, Lauridsen, Ostengaard and Qvortrup (2023) conduct a systematic critical

review of the validity evidence for a widely used questionnaire and find that its use has run ahead of that evidence.

Both cases illustrate the asymmetry this article addresses. Constructing a measure attracts methodological scrutiny; deriving a statistic from measures already constructed usually does not, even though the derived statistic is what enters the argument.

2.6 The research gap

No published analysis of Indonesian provincial education data reports a rank correlation or an influence diagnostic alongside a Pearson coefficient. This article computes both for four relationships in the same data, identifies the one that fails, and quantifies how far it fails.

3. METHOD

Three Central Statistics Agency tables are used, all retrieved on 30 August 2026: Angka Partisipasi Sekolah (APS) Menurut Provinsi dan Kelompok Umur for 2025, covering thirty-eight provinces and four age bands; Tingkat Penyelesaian Pendidikan Menurut Jenjang Pendidikan dan Provinsi for 2023, covering thirty-four provinces with data at three levels; and Indeks Pembangunan Literasi Masyarakat dan Dimensi Penyusunnya Menurut Provinsi for 2025, covering thirty-three provinces with data on a composite and its two dimensions.

Four bivariate relationships are examined. Two are within the participation table: ages sixteen to eighteen against ages nineteen to twenty-three, and ages seven to twelve against ages nineteen to twenty-three. One is within the completion table: primary against upper secondary completion. One is within the literacy table: the compliance dimension against the performance dimension.

For each relationship two coefficients are computed across the provinces with data: the Pearson product-moment correlation, and the Spearman rank correlation, computed as the Pearson correlation of the ranks. The gap between them is reported. For the relationship that fails, a leave-one-out analysis recomputes the Pearson coefficient thirty-eight times, each time omitting one province, and reports the provinces whose omission moves it most.

No observation is deleted from any reported headline figure. The leave-one-out results are diagnostic and are presented as such; Section 4.6 argues against deletion as a remedy.

Five limitations apply. First, the Pearson-Spearman gap is a heuristic and not a test; it has no critical value and this article proposes none. Second, a large gap can arise from a genuine non-linear but monotone relationship as well as from outliers, and the leave-one-out analysis is what distinguishes the two. Third, the three tables cover different years and different numbers of provinces, so the four relationships are not computed on a common population. Fourth, one year is observed for each. Fifth, the analysis is descriptive.

4. RESULTS AND DISCUSSION

4.1 Four relationships, two coefficients

Table 1 gives both coefficients for each of the four relationships.

Table 1. Pearson and Spearman correlations for four relationships in Indonesian education data

Computed by the authors from Central Statistics Agency tables on school participation by province and age group (2025), education completion by level and province (2023), and the community literacy development index and its dimensions (2025), all retrieved 30 August 2026. Units are the provinces with data for both variables in the pair.

Three of the four relationships show a gap between 0.0655 and 0.0722, and in all three the Spearman coefficient is the smaller. A gap of this size in this direction is unremarkable: the Pearson coefficient uses the values and the Spearman only the order, and where the relationship is broadly linear the two agree closely.

The fourth relationship has a gap of 0.4159, six times the largest of the others. Its Spearman coefficient of 0.0734 is close to zero: among the thirty-eight provinces, the ordering by primary participation and the ordering by participation at ages nineteen to twenty-three are almost unrelated.

The three well-behaved relationships establish that the diagnostic is not simply flagging everything. Data drawn from the same statistical system, at the same administrative level, on comparable scales, produce small gaps in three cases and a very large one in the fourth.

4.2 Locating the influence

Table 2 reports the leave-one-out analysis for the relationship that fails.

Table 2. Effect on the Pearson coefficient of omitting a single province, participation 7–12 against 19–23

Computed by the authors from the Central Statistics Agency participation table named at Table 1. Each row reports the Pearson coefficient recomputed over the remaining 37 provinces. Provinces are shown in order of the absolute size of the change, except the first row.

Papua Pegunungan and Papua Tengah are the two provinces whose omission reduces the coefficient, and they reduce it far more than any other province changes it. Both sit at the bottom of both variables: 82.24 and 80.13 per cent at ages seven to twelve, against a mean of 98.06, and 12.48 and 15.48 per cent at ages nineteen to twenty-three, against a mean of 29.48.

DI Yogyakarta has the largest effect in the opposite direction, its omission raising the coefficient by 0.0704. It sits at the top of both variables and is therefore also on the line the two Papua provinces define, but as a single high point among a dense cluster its leverage is smaller.

The mechanism is standard. Two observations far below the mass of the data, low on both variables, create a linear association between two variables that are otherwise unrelated. The remaining thirty-six provinces occupy a narrow band on the first variable — thirty-three of them within two percentage points of one another — within which the second variable varies freely.

4.3 What remains when the two are removed

Table 3 recomputes the four relationships without the two highland provinces, where they are present in the data.

Table 3. Coefficients with and without Papua Tengah and Papua Pegunungan

Computed by the authors from the three Central Statistics Agency tables named at Table 1. The two highland provinces are absent from the completion and literacy tables, so those rows omit Papua instead for completion, and the five lowest-scoring provinces

for literacy, as the nearest available equivalent; the literacy Spearman without is not reported because the exclusion set is not comparable.

The first row is the article's central result. A Pearson coefficient of 0.4893 becomes 0.0178 on removal of two observations from thirty-eight — a reduction of 96 per cent. The Spearman coefficient moves from 0.0734 to -0.0896 , both of which are indistinguishable from no relationship.

The second row shows what a robust relationship looks like under the same operation. The coefficient falls from 0.7337 to 0.6323, a reduction of 14 per cent, and the Spearman coefficient from 0.6615 to 0.6021. The association between participation at ages sixteen to eighteen and at nineteen to twenty-three survives the removal of the extremes because it holds among the ordinary provinces too.

The third and fourth rows behave similarly. Primary against upper secondary completion is 0.6725 across thirty-four provinces and 0.6620 without Papua, the lowest of them. The two literacy dimensions correlate at 0.9462 across thirty-three provinces and 0.8545 among the twenty-eight after removing the five lowest.

4.4 The same failure in a different pair

The pattern is not confined to the participation table, and a second instance from a different pairing strengthens the general point.

A companion analysis relates participation at ages sixteen to eighteen to a supply measure, the ratio of villages in a province with a primary school to villages with an upper secondary school. Across all thirty-eight provinces the Pearson correlation is -0.6039 and the Spearman -0.3791 , a gap of 0.2248 — smaller than the 0.4159 above but well beyond the 0.07 typical of the well-behaved pairs.

Removing the two highland provinces takes the Pearson coefficient to -0.4419 and the Spearman to -0.2708 . The relationship survives in reduced form, which is a third case: neither robust nor artefactual, but partly generated by the extremes.

The supply variable has its own extreme, Papua Selatan at 14.32 primary-school villages per upper-secondary-school village against a national ratio of 3.92. Removing that province alone strengthens the Pearson coefficient to -0.6482 , so the influential observations do not all pull in the same direction and a single leave-one-out pass is needed to see which do.

The three cases together give a small typology. Where the gap is under about 0.1, the relationship holds throughout the data. Where it is around 0.2, the extremes contribute materially but the relationship survives. Where it approaches 0.4, the relationship may be entirely theirs.

4.5 Why the primary variable behaves this way

The failure in the first row of Table 1 is not an accident of these particular data, and its cause is worth stating because it will recur wherever a measure sits near its bound.

Participation at ages seven to twelve is bounded above by 100 per cent, and thirty-three of the thirty-eight Indonesian provinces report at least 98 per cent. The variable therefore has almost no variance to contribute among those thirty-three, and any correlation involving it must be produced by the five provinces below the cluster.

This is the same structural situation Goldhaber, Grout, Wolff and Martinkova (2021) identify in reference ratings that cluster near the top of their scale, and that Min and Aryadoust (2021) discuss for compressed assessment measures. A variable at its ceiling

can still enter a regression or a correlation and will produce a number; the number will describe the units off the ceiling.

The practical rule that follows is narrow and useful. Before correlating a bounded educational measure with anything, count how many units sit within a small distance of the bound. Where the answer is most of them, the correlation is about the remainder, and it should be reported as such.

The second variable in the failing pair has no such problem: participation at ages nineteen to twenty-three ranges from 12.48 to 54.72 per cent and uses its scale fully. The failure is attributable to one side of the pair.

4.6 Against deletion as a remedy

Having established that two provinces produce the coefficient, the natural next step is to delete them and report the remainder. This article argues against that and the reasons are worth setting out.

Papua Tengah and Papua Pegunungan are not measurement errors. They are provinces of the Republic of Indonesia, created in 2022, with populations of children whose participation rates of 80.13 and 82.24 per cent at ages seven to twelve are the most consequential educational facts in the table. An analysis that removes them to obtain a cleaner coefficient has removed its own subject.

The alternative is to report both coefficients, the leave-one-out result, and the substantive description of the influential units — which is what Tables 1 to 3 do. The reader then knows that primary participation does not predict later participation among the thirty-six ordinary provinces, and separately that two provinces are far below the rest on both. Both statements are true and neither is available from the coefficient alone.

Prentice (2022) and Jover, Manzano, Llos and Termes (2025) both study populations defined by their distance from the norm, and in both the tail is the point. Serrano-Ardila, Baena-Navarro and Carriazo-Regino (2026) address rural education specifically for the same reason. The methodological posture recommended here is consistent with that literature: describe the tail, do not discard it.

Where deletion is genuinely appropriate is where the observation is an error, a different kind of unit, or outside the population of interest. None of those applies here.

4.7 Implications for practice

Three implications follow.

First, report both coefficients. The Spearman correlation costs one line of code and its divergence from the Pearson value is a sufficient signal to look further. Had the pair 0.4893 and 0.0734 been published together, no reader would have taken the first as a description of Indonesian provinces.

Second, run leave-one-out on any correlation that will carry weight. With thirty-eight units this is thirty-eight recomputations and it takes no appreciable time. The output is a short list of the observations the finding depends on, which is information a reader is entitled to.

Third, treat a bounded measure with most units near the bound as unavailable for correlation without explicit comment. The count of units near the bound should be reported alongside the coefficient. Fazlul, Koedel and Parsons (2023) make the parallel argument for a proxy that does not measure its concept: the remedy is disclosure, not silence.

A fourth point concerns cumulative practice. Where a literature reports correlations without these diagnostics, individual failures are invisible and may propagate through meta-analysis and policy. Bokhove, Jerrim and Sims (2024) show the value of interrogating an official measure directly rather than accepting it; the same applies to a derived one.

4.8 Limitations of the finding

Four limitations bound the result.

The Pearson-Spearman gap is a heuristic. This article offers no critical value and none should be inferred from the three-case typology in Section 4.4, which rests on four relationships in one country. The gap indicates where to look and does not by itself establish that a relationship is artefactual; the leave-one-out analysis is what does that.

A large gap can also arise from a genuine monotone non-linearity, in which case the Spearman coefficient is the better summary and the Pearson coefficient is the one to distrust. Distinguishing this from the outlier case requires the leave-one-out pass, and this article did not encounter an instance.

The four relationships are computed on three tables with different coverage — thirty-eight, thirty-four and thirty-three provinces — and different reference years. They are not a controlled comparison, and the well-behaved cases could differ from the failing one for reasons besides the presence of extremes.

Finally, this is one country in one period. The mechanism identified — a bounded measure near its ceiling, with a small number of units far below — is general, but the frequency with which it occurs in educational data elsewhere is unknown and would be worth establishing.

5. CONCLUSION

The Pearson correlation between Indonesian provincial school participation at ages seven to twelve and at ages nineteen to twenty-three is 0.4893. The Spearman rank correlation is 0.0734, and the Pearson coefficient recomputed without Papua Tengah and Papua Pegunungan is 0.0178.

Two provinces of thirty-eight therefore produce 96 per cent of a coefficient that would ordinarily be read as evidence that provinces performing well in primary schooling perform well later. Among the remaining thirty-six provinces there is no such relationship.

The diagnostic that reveals this is the gap between the two coefficients. Three other relationships in the same body of data show gaps of 0.0655, 0.0658 and 0.0722 and survive the removal of their extremes: the association between participation at ages sixteen to eighteen and nineteen to twenty-three falls only from 0.7337 to 0.6323, and primary against upper secondary completion from 0.6725 to 0.6620. The failing relationship has a gap of 0.4159.

The cause is on one side of the pair. Participation at ages seven to twelve is bounded above at 100 per cent and thirty-three of thirty-eight provinces report at least 98 per cent, so the variable has almost no variance to contribute except through the five provinces below the cluster.

The remedy is not deletion. Papua Tengah and Papua Pegunungan are the least-served jurisdictions in the country and an analysis that removes them to obtain a cleaner number has discarded what it exists to study. The recommendation is to report both coefficients, to run and publish a leave-one-out pass on any correlation that will carry

weight, and to state how many units sit near the bound of a bounded measure before correlating it with anything.

REFERENCES

- Adeleke, O., & McSharry, P. E. (2022). Female enrollment, child mortality and corruption are good predictors of a country's UN Education Index. *International Journal of Educational Development*, 90, 102561. <https://doi.org/10.1016/j.ijedudev.2022.102561>
- Ahmed, M. (2023). Education in perennial crisis: have we been asking the right questions?. *International Journal of Educational Development*, 103, 102910. <https://doi.org/10.1016/j.ijedudev.2023.102910>
- Armand, M. K., Herman, P. R., & Honore, T. O. (2024). Inequality of educational opportunity among primary school children in Cameroon. *International Journal of Educational Development*, 109, 103103. <https://doi.org/10.1016/j.ijedudev.2024.103103>
- Badan Pusat Statistik. (2026). Angka partisipasi sekolah (APS) menurut provinsi dan kelompok umur, 2025 [School participation rate by province and age group, 2025]. Retrieved August 30, 2026, from <https://www.bps.go.id/id/statistics-table/2/MjIxMSMy/angka-partisipasi-sekolah--aps--menurut-provinsi-dan-kelompok-umur.html>
- Badan Pusat Statistik. (2026). Tingkat penyelesaian pendidikan menurut jenjang pendidikan dan provinsi, 2023 [Education completion rate by level and province, 2023]. Retrieved August 30, 2026, from <https://www.bps.go.id/id/statistics-table/2/MTk4MCMY/tingkat-penyelesaian-pendidikan-menurut-jenjang-pendidikan-dan-provinsi.html>
- Bokhove, C., Jerrim, J., & Sims, S. (2024). Do schools that employ an inspector get better inspection grades?. *British Educational Research Journal*, 50(5), 2383-2406. <https://doi.org/10.1002/berj.4025>
- Coates, A. (2026). A diagnostic classification model of TIMSS 2023 grade four mathematics. *Studies in Educational Evaluation*, 89, 101615. <https://doi.org/10.1016/j.stueduc.2026.101615>
- Dearing, E., Zachrisson, H. D., Bustamante, A. S., & Vandell, D. L. (2026). Developmental timing of educational opportunities for children in low-income households. *Educational Researcher*. <https://doi.org/10.3102/0013189x261463238>
- Doz, D., Bone, J., & Felda, D. (2026). Towards a multidimensional grade in mathematics in Slovenia: a fuzzy logic approach. *Studies in Educational Evaluation*, 89, 101595. <https://doi.org/10.1016/j.stueduc.2026.101595>
- Engel, L., & Burch, P. (2021). Policy sociology in the contemporary global era: continued importance and pressing methodological considerations. *Educational Researcher*, 50(7), 474-478. <https://doi.org/10.3102/0013189x211009184>
- Fazlul, I., Koedel, C., & Parsons, E. (2023). Free and reduced-price meal enrollment does not measure student poverty: evidence and policy significance. *Economics of Education Review*, 94, 102374. <https://doi.org/10.1016/j.econedurev.2023.102374>
- Ford, K. S., Iantosca, M. H., & Cate, L. (2025). Contesting racial categories: variability and complexity in university student ethnoracial self-identifications. *Educational Researcher*, 54(3), 132-140. <https://doi.org/10.3102/0013189x241310170>
- Gates, E. F., Teasdale, R. M., Shim, C., & Hubacz, H. (2024). Whose and what values? Advancing and illustrating explicit specification of evaluative criteria in education.

- Studies in Educational Evaluation, 81, 101335. <https://doi.org/10.1016/j.stueduc.2024.101335>
- Goldhaber, D., Grout, C., Wolff, M., & Martinkova, P. (2021). Evidence on the dimensionality and reliability of professional references' ratings of teacher applicants. *Economics of Education Review*, 83, 102130. <https://doi.org/10.1016/j.econedurev.2021.102130>
- Jiang, Y., Ko-Wong, L., & Valdovinos Gutierrez, I. (2025). The feasibility and comparability of using artificial intelligence for qualitative data analysis in equity-focused research. *Educational Researcher*, 54(3), 153-163. <https://doi.org/10.3102/0013189x251314821>
- Jover, A., Manzano, M., Llos, B., & Termes, A. (2025). Emotional experiences and stigma among families benefiting from Barcelona's shock plan against school segregation and for inclusion and equal opportunities. *British Educational Research Journal*, 52(2), 1159-1177. <https://doi.org/10.1002/berj.70054>
- Kelchen, R., Ortagus, J., Rosinger, K., Baker, D., & Lingo, M. (2023). The relationships between state higher education funding strategies and college access and success. *Educational Researcher*, 53(2), 100-110. <https://doi.org/10.3102/0013189x231208964>
- Luschei, T. F., Chudgar, A., Kendall, N., Chirkina, T., Choi, S. Y., Gidra, A., Chavarria, Y., Chimwaza, T. M., Isaac, A. T., Makhuva, S., Muwa, I. T., & Thangaraj, M. S. (2026). Student well-being, sense of belonging, and alienation among secondary school students in Colombia, India, and Malawi. *International Journal of Educational Development*, 124, 103628. <https://doi.org/10.1016/j.ijedudev.2026.103628>
- Min, S., & Aryadoust, V. (2021). A systematic review of item response theory in language assessment: implications for the dimensionality of language ability. *Studies in Educational Evaluation*, 68, 100963. <https://doi.org/10.1016/j.stueduc.2020.100963>
- O'Neill, L., Lauridsen, H. H., Ostengaard, L., & Qvortrup, A. (2023). Validity evidence for the Experiences of Teaching and Learning Questionnaire in evaluations of quality learning: a systematic critical literature review. *Studies in Educational Evaluation*, 78, 101283. <https://doi.org/10.1016/j.stueduc.2023.101283>
- Prentice, C. M. (2022). Educators' positive practices with refugee pupils at two schools in England. *British Educational Research Journal*, 48(6), 1125-1144. <https://doi.org/10.1002/berj.3818>
- Sanfo, J. B. M., & Malgoubri, I. (2023). Teaching quality and student learning achievements in Ethiopian primary education: how effective is instructional quality in closing socioeconomic learning achievement inequalities?. *International Journal of Educational Development*, 99, 102759. <https://doi.org/10.1016/j.ijedudev.2023.102759>
- Serrano-Ardila, L., Baena-Navarro, R., & Carriazo-Regino, Y. (2026). Digital integration in rural education: foundations for a sustainable pedagogical model. *International Journal of Educational Development*, 121, 103516. <https://doi.org/10.1016/j.ijedudev.2026.103516>
- Stals, L., Isac, M. M., & Claes, E. (2024). Political trust among European youth: evaluating multi-dimensionality and cross-national measurement comparability. *Studies in Educational Evaluation*, 80, 101321. <https://doi.org/10.1016/j.stueduc.2023.101321>
- Undang-Undang Nomor 16 Tahun 1997 tentang Statistik [Law No. 16 of 1997 on Statistics].

Viano, S., Baker, D. J., Ford, K. S., & Johnston-Guerrero, M. P. (2024). A latent class analysis of racial terminology in education research: patterns of racial classifications in AERA journals. *Educational Researcher*, 53(4), 213-222. <https://doi.org/10.3102/0013189x241227901>

Wagner, D. A., Nair, T., Thapa, A., & Kumar, A. (2025). The Gini Learning Index: a new framework to measure learning inequality across contexts. *International Journal of Educational Development*, 119, 103433. <https://doi.org/10.1016/j.ijedudev.2025.103433>

White, M., & Maher, B. (2024). What value do standardized observation systems add to summative teacher evaluation systems?. *Studies in Educational Evaluation*, 83, 101414. <https://doi.org/10.1016/j.stueduc.2024.101414>

Yan, Z., & Pastore, S. (2022). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation*, 74, 101183. <https://doi.org/10.1016/j.stueduc.2022.101183>

DECLARATIONS

Ethics statement. Aggregate secondary data from public agencies only. No human participants, no identifiable personal data, no animal subjects; ethics approval not required.

Peer review. Double-anonymous.

6. ARTICLE INFORMATION

Copyright and licence. © 2026 The Author(s). Open access under CC BY-SA 4.0: <https://creativecommons.org/licenses/by-sa/4.0/>

Article DOI. [to be assigned on publication]

How to cite. Author, A., Author, B., & Author, C. (2026). Two provinces of thirty-eight: when a correlation is an artefact of its outliers. *Journal of Education, Linguistics and Psychology*, 2(1), [pages]. [https://doi.org/\[to be assigned\]](https://doi.org/[to be assigned])

Publisher. Research Center Progresif, Indonesia · jelp@researchcenterprogresif.org

© 2026 The Author(s). Published by Research Center Progresif, Semarang City, Central Java, Indonesia.

This is an open access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International Licence (CC BY-SA 4.0), which permits use, sharing and adaptation provided the original work is properly cited.

How to cite: Yahya Syahrizal (2026). TWO PROVINCES OF THIRTY-EIGHT: WHEN A CORRELATION IS AN ARTEFACT OF ITS OUTLIERS. JELP, 1(1), 12-22.