methods, data, analyses | 2026, pp. 1-30 DOI: 10.25521/mda.578

Let's Measure Concentration and Dispersion in Ordinal Data

Clem Aeppli[1], Didier Ruedin[2,3]

[1] *Harvard University*
[2] *University of Neuchâtel*
[3] *University of the Witwatersrand*

Abstract

In the social sciences, we often use ordinal data, like 'agree' to 'disagree'. Frequently, we also want to know whether answers are clustered, whether they are polarized, etc. While many measures have been designed to capture what is variably referred to as agreement, consensus, concentration, or dispersion and polarization in ordinal data, little is known how they differ in practice—including how they react to changes in the distribution of responses. To better understand how different measures work in practice, we compare consensus scores across specific situations: constructed cases, simulated data where we know the underlying distribution, and empirical data. We highlight the similarities as well as the distinctive features of different approaches that have been developed across a disparate literature. In most cases, the different measures are highly correlated, but there are cases where the choice of method leads to substantially different conclusions. These situations are difficult for non-mathematical users to predict. We recommend a combination of measures for robustness and graphs to examine the distribution qualitatively.

Keywords: agreement, consensus, polarization, ordinal data, measurement

© The Author(s) 2026. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

In the social sciences, we often encounter ordinal data—just think of Likert survey items ranging from 'very strongly agree' to 'very strongly disagree,' survey ratings, customer feedback in the form of star ratings, or ordinal categorical data produced in content analysis when coding the position of a text. To give an idea of the range of applications, Krymkowski et al. (2009) sought to capture the degree of 'norm crystallization' among outdoor recreators; Blair and Lacy (2000) compared the degree of consensus over belief in God between Black and White Americans; Haas and Hüllermeier (2025) assessed goodwill claims of car manufacturers; Harding (2011) and Berg et al. (2013) examined how cultural heterogeneity affects educational attainment; or Reardon (2009) studied occupational segregation. Indeed, disagreement and consensus of public opinion more generally are mainstays of the social sciences (e.g., Balietti et al., 2021), and scholars regularly want to know to what extent positions are concentrated or dispersed and assess the degree of ideological polarization from Likert-type survey responses or survey ratings using dispersion (e.g., Alwin & Tufiş, 2016; Enders, 2021). The need to accurately measure dispersion and disagreement has become even more pressing in an era marked by political polarization (Kubin & von Sikorski, 2021; Levendusky, 2009; Lupu, 2015), but there is little guidance available on how to measure these phenomena adequately.

In particular, it remains unknown how the different measures react to different distributions, and researchers have little guidance about which approach to follow. This contrasts with the situation for continuous data, where the sample standard deviation can be used to measure concentration or dispersion. When faced with ordinal data, however, researchers often hesitate to use the standard deviation: If we use the sample standard deviation for ordinal data, we assume that the categories are equally spaced, an assumption that is often violated or undetermined.[1] In a disparate literature spanning several disciplines, researchers have developed different measures of ordinal concentration and dispersion. The range of terminology employed demonstrates that such measures are often developed in isolation from one another. Various measures refer to 'consensus' (Davies, 1970; Leik, 1966; Tastle & Wierman, 2007), 'concentration' (Blair & Lacy, 2000), or 'agreement' (Van der Eijk, 2001) to describe the clustering of responses.

[1] Another assumption is that the data come from a probability sample, which is arguably also often ignored. This article will assume that observations are chosen with equal probability and do not need to be weighted; hence, the focus of this article is on practical application and not on developing a proper inferential and asymptotic theory. For some of the measures we discuss, sampling weights could be envisaged.

Direct correspondence to
Didier Ruedin, University of Neuchâtel, Rue Abram-Louis-Breguet 2,
2000 Neuchâtel, Switzerland
E-mail: didier.ruedin@unine.ch

This diversity of approaches leaves researchers unsure how to proceed, worried about introducing biases in research. We need to know which of the different approaches is the most appropriate in specific research applications, and to do so, we need to understand how they behave and how bad those violations of assumptions are in practice in the case of the sample standard deviation. The aim of this article is to provide a comprehensive introduction to existing measures and especially to compare these measures to provide researchers working with ordinal data with practical guidance.

We present different measures and show how they react differently to data modifications. We use one pair of terms: 'concentration' and its complementary concept 'dispersion.'[2] Concentration exists when many responses cluster around a particular point, whereas dispersion occurs when responses are spread out or scattered. Our choice highlights the fact that ordinal data need not reflect the opinions of respondents but may refer to positions more generally. To better understand how different measures behave in practice, we provide axiomatic scenarios, simulate ordinal data from latent continuous variables, and examine real-world examples. In most cases, the different measures are highly correlated, but we show that two groups of measures exist that can lead to different results. The differences are difficult to anticipate because they depend on the precise coding of ordinal data. However, the contrived examples and simulations allow us to identify patterns that may guide researchers in choosing a particular measure. In the absence of prioritization by researchers, we recommend a combination of measures for robustness and highlight the benefit of qualitatively assessing distributions using graphs.

Measurement of Concentration and Dispersion

As we often do in the social sciences, researchers can treat an ordinal variable as continuous (known as pseudo-interval or pseudo-continuous; Bartholomew et al., 2008, p. 245), and we can use the sample standard deviation or other measures such as the interquartile range to measure concentration and dispersion. This is particularly the case for Likert items which are routinely treated as continuous, although not all researchers agree. When doing so, we assume that categories are equally spaced, and by using the standard deviation, researchers (implicitly) defend this assumption as reasonable. Equal spacing means that the response categories are evenly spaced in a substantive sense: that category 1 is

[2] We realize that the term 'dispersion' is sometimes used to mean variance, but so are alternative terms such as 'spread' or 'variability'. By choosing this pair of terms, we clarify how the terms are used in this article, and we do not seek to prescribe terminology on other researchers, acknowledging how these and related terms are used differently across the social sciences.

as far from category 2 as category 2 is from category 3. Because we know that this assumption does not necessarily hold up or because typically we cannot test it, researchers often hesitate. For example, Blair and Lacy (2000) argue that the 'true' value of categories may follow a logarithmic pattern with scores of ln(1), ln(2), and ln(3) instead of 1, 2, and 3. Under this assumption, the categories are substantively 'closer' together as they increase in the index, and the categorical standard deviation can give misleading or indirect answers (see also Anderson, 2025 on unequal spacing in the case of self-assessed social status). Furthermore, Van der Eijk (2001) notes that if the mean of an ordinal distribution is close to either of the endpoints of the scale, the standard deviation is greatly influenced by additional points at the other extreme of the scale, even if there are few such points. In response to these concerns, alternative measures have been introduced (Table 1).

When developing alternative measures of ordinal concentration, some scholars first identify the point of maximal concentration as any distribution in which all responses fall into a single category (Leik, 1966; Van der Eijk, 2001). Maximal dispersion is defined as the distribution in which half of the responses are in each extreme category, a distribution some researchers refer to as polarization. Blair and Lacy (2000), Kvålseth (1988), Leik (1966) and others have also noted that a measure of ordinal concentration should equal 1 at the point of maximal concentration and 0 at the point of maximal dispersion, a characteristic the standard deviation does not have. This standardization allows comparisons, and a measure of ordinal dispersion can be defined as 1 less the measure of concentration, ranging from 0 at the point of maximal concentration to 1 at the point of maximal dispersion. Considering this possibility of converting between the two, in the following we focus on concentration.

Table 1

Overview of Different Measures Compared in the Article

Measure	References	Notes
A	Van der Eijk (2001)	
L1	Leik (1966)	
L2	Blair and Lacy's (2000) /	Kvålseth's (1995) COV is 1 – *L2*
$L2^2$	Blair and Lacy's (2000) /^{2}; Reardon (2009)	Berry and Mielke's (1992) IOV is 1 – *L2*
Cns	Tastle and Wierman (2007)	
SD		
ln(*SD*)		

Note. All measures take a frequency vector as the argument: a vector with the number of observations in each ordinal category. For instance, a frequency vector (10,12,3,4,2) may represent 10 'strongly disagree', 12 'disagree', …, and 2 'strongly agree'.

Measures take the form of algorithms, as in Van der Eijk's (2001) measure of agreement *A,* or of formulas. The algorithm for agreement *A* breaks the vector of responses into uniform layers. These layers are defined by their pattern of contiguous non-empty and empty categories and by the number of observations in non-empty categories. Each layer is assigned an agreement value based on the length of empty categories in the layer. The sum of these agreement values, weighted by the size of the layer, is the resulting level of agreement (Appendix A provides more detail on Van der Eijk's (2001) algorithm, including a worked example).

There are two common methods by which researchers develop formulas to measure ordinal concentration, attempting to maintain the assumptions of ordinal data while avoiding additional assumptions of distance or category width. One approach uses the empirical cumulative mass function (Blair & Lacy, 2000; Kvålseth, 1995; Leik, 1966). A different approach is to draw on the Shannon entropy to calculate the distance between each observation and the mean (Tastle and Wierman's (2007) measure of consensus, *Cns*). In the following, we denote the number of response categories—the length of the response vector—by *L*. For instance, a Likert item with 'strongly disagree', 'disagree', 'neither', 'agree', and 'strongly agree' has a length of 5. We refer to the total number of responses as *R*; the number of responses in each category *i* is referred to by r_i, and the fraction of total responses in category *i* by p_i. The empirical cumulative mass function $F = \{F_1, F_2, \ldots, F_L\}$ that is often invoked by such formulas is defined by $F_i := \sum_{j \le i} \frac{r_j}{R} = \sum_{j \le i} p_j$. Many of the measures proposed in the literature are specific instances of a generalized family of measures that we call the L_N family. Leik's (1966) *D* corresponds to *L1* and Blair and Lacy's (2000) *l* corresponds to *L2* while Blair and Lacy's (2000) l^2 corresponds to L_2^2.

Technically speaking, Blair and Lacy (2000) define $F^{maxdisp} = (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, 1)$, the cumulative mass function corresponding to the point of minimal concentration and maximal dispersion in which the responses are split equally between each extreme. Similarly, we can define $F^{maxconc} = (1,1,1,\ldots,1)$ as a cumulative mass function corresponding to one of the points of maximal concentration where all responses fall into a single category. Blair and Lacy (2000) observe that Leik's (1966) *D,* Berry and Mielke's (1992) *IOV,* Kvålseth's (1995) *COV,* and their (2000) *l* and l^2, all belong to a family of measures of the form

$$\text{concentration} = \frac{d(F_{-L}, F_{-L}^{maxdisp})}{d(F_{-L}^{maxconc}, F_{-L}^{maxdisp})},$$

where *d* is a distance function defined on the first *L*-1 entries of each cumulative mass vector. We have added the subscript -*L* to *F*, $F^{maxdisp}$ and $F^{maxconc}$ to make clear that we are only measuring the first *L*-1 entries in these vectors. Dividing by $d(F_{-L}^{maxconc}, F_{-L}^{maxdisp})$ ensures that the resulting index is standardized between

0 and 1. Zero corresponds to maximum dispersion, and 1 corresponds to maximum concentration. The complementary measure dispersion = 1 – concentration follows from this definition. Note that, while there are L possible maximally concentrated cumulative mass functions $F^{maxconc}$—corresponding to the L possible positions in which all mass can be placed—for each of them, the denominator $d(F_{\cdot L}^{maxconc}, F_{\cdot L}^{maxdisp})$ will be the same under any of the distance functions d discussed below.

The precise nature of the distance function, d, varies according to the measure used. Leik's measure uses the *L1* distance, commonly known as the 'city-block' distance. This means that Leik's (1966) D can be written as:

$$D = \frac{L1\,(F_{\cdot L}, F_{\cdot L}^{maxdisp})}{L1(F_{\cdot L}^{maxconc}, F_{\cdot L}^{maxdisp})} = \frac{\sum_{i=1}^{L-1} |F_i - \frac{1}{2}|}{\frac{(L-1)}{2}}.$$

Blair and Lacy (2000) define a measure l of concentration that uses the *L2* distance. Thus, l is written as:

$$l = \frac{L2\,(F_{\cdot L}, F_{\cdot L}^{maxdisp})}{L2(F_{\cdot L}^{maxconc}, F_{\cdot L}^{maxdisp})} = \left[\frac{\sum_{i=1}^{L-1} (F_i - \frac{1}{2})^2}{\frac{(L-1)}{4}}\right]^{\frac{1}{2}}.$$

Berry and Mielke's (1992) *IOV* measure of dispersion is equal to $1–l^2$, while Kvålseth's (1995) *COV* measure of dispersion is equal to $1–l$.

The second approach expands on the Shannon entropy formula. Tastle and Wierman (2007) define a measure of consensus *Cns* that does not explicitly rely on the cumulative-distribution approach. Instead, their measure incorporates the sample mean $\overline{X}$ of the responses X_i:

$$Cns(X_1, X_2, \ldots, X_R) = 1 + \sum_{i=1}^{R} p_i \log_2 (1 - \frac{|X_i - \overline{X}|}{d_X}).$$

Here, $d_x := \max(X) – \min(X)$ is the width of the categories (Tastle & Wierman, 2007). A question with response categories 1, 2, and 3 has a length of 3–1=2. Although it does not invoke the cumulative distribution function, Tastle and Wierman's (2007) *Cns* measure relies on $\overline{X}$, which treats indices of categories as absolute positions with equal distance between them.

Davies (1970) remarked that because it combines linearly the cumulative distribution function F, Leik's measure D assumes that the categories are equally spaced. It is possible that such "supra-ordinal assumptions" (Davies 1970, 27) affect the validity of other measures using F. For these reasons, we compare all measures to the standard deviation calculated over the category indices. Measures that are indistinguishable from the categorical standard deviation may

indeed make supra-ordinal assumptions; moreover, given the choice between a supposedly ordinal measure and the categorical standard deviation, it makes sense to choose the simpler one if both formulas give comparable results.

For some of the measures, confidence intervals can be derived analytically. Blair and Lacy (2000) offer analytical measures of variance of the *l* index for both small and large samples, permitting the construction of analytical confidence intervals. Despite this, hypothesis testing remains difficult to conduct. Blair and Lacy note that the null hypothesis that $l = r$, for example, would describe many points with *L2* distance *r* from $F_{\cdot L}^{maxdisp}$, a problem that increases in difficulty as the number of response categories increases (Blair & Lacy, 2000). Attempting a hypothesis test with Leik's *D* results in the same difficulty. Tastle and Wierman (2007) do not obtain an analytical measure of variance for *Cns*, nor does Van der Eijk (2001) for the measure *A*. To overcome the insufficient development of analytical measurements of variance of these indices, we use the bootstrap method to see whether the differences in concentration between different values of dispersion or concentration are statistically reliable.

A different class of measures exists in indices of qualitative variation (*IQV*, Berry & Mielke, 1992) developed to assess dispersion in nominal data. These are unsuitable for ordinal data because, by design, they fail to consider the important ordered aspect of data: a concentration in the categories 'strongly agree' and 'neutral' would be indistinguishable from a concentration in the categories 'strongly agree' and 'strongly disagree' (see also DiMaggio et al., 1996). A five-category distribution of (0, 30, 40, 30, 0), with a clear concentration at the center, receives the same *IQV* score as the dispersed distribution (30, 30, 0, 0, 40), which is undesirable with ordinal data. Similarly, measures of intercoder agreement capture the similarity between different sets of responses or codes, not whether positions within a distribution do so—though in this case, ordinal data also complicates matters a bit (Tran et al., 2020).

In the analysis, we compare Leik's *D* (*L1*), Blair and Lacy's *l* (*L2*), the squared variant $L2^2$, Tastle and Wierman's *Cns*, Van der Eijk's *A*, and the categorical standard deviation. We compare the different measures in three ways: First, we use contrived examples proposed in the literature. Second, because there is no intuitive way to determine which measures are more accurate, we suppose that ordinal responses correspond to an underlying, continuous latent variable. Variants of this assumption are central in many ordinal models, such as the ordinal logistic regression model. Third, we use real data to explore the implications in actual research. In all simulations we linearly transform *A* so that it ranges between 0 and 1, with 1 corresponding to the maximum concentration. We omit direct derivatives because they do not affect the substantive conclusions.

Contrived Examples by Tastle and Wierman

To evaluate their measure of consensus *Cns*, Tastle and Wierman (2007, p. 7) provided several sets of ideal typical distributions. These cases help us to understand the dynamics of the different measures where there is progression between cases, but they are also useful for understanding cases close to the extremes. We use these to gauge how different measures react to different distributions. Readers may decide themselves whether to imagine a swing in political opinion, polarization, or other changes in dispersion. The corresponding concentration scores are presented in Table 2. These contrived examples address the concern that the continuous standard deviation does not always accurately capture what

Table 2

Concentration Scores for Ideal-Typical Distributions, Including Those Shown in Figure 1

Distribution	*A*	*L1*	*L2*	$L2^2$	*Cns*	*SD*	ln(*SD*)
A	0.000	0.000	0.000	0.000	0.000	2.010	0.809
B	0.431	0.250	0.354	0.125	0.292	1.589	0.634
C	0.500	0.400	0.447	0.200	0.434	1.421	0.571
D	0.603	0.485	0.539	0.291	0.518	1.276	0.539
E	0.583	0.500	0.707	0.500	0.585	1.005	0.348
F	0.675	0.600	0.632	0.400	0.634	1.101	0.437
G	0.712	0.700	0.700	0.490	0.700	1.101	0.463
H	0.875	0.750	0.866	0.750	0.807	0.503	0.348
I	0.875	0.750	0.866	0.750	0.807	0.503	0.204
J	1.000	1.000	1.000	1.000	1.000	0.000	0.000
K	0.000	0.000	0.000	0.000	0.000	2.108	0.848
Q	0.292	0.250	0.500	0.250	0.322	1.581	0.731
R	0.583	0.500	0.707	0.500	0.585	1.054	0.579
S	0.875	0.750	0.866	0.750	0.807	0.527	0.365
T	1.000	1.000	1.000	1.000	1.000	0.000	0.000
U	0.792	0.750	0.791	0.625	0.686	1.000	0.347

Note. Standard deviations were calculated under the assumption that the distributions refer to values 1, 2, 3, 4, and 5, respectively. For distributions A to J, we assumed 100 observations, as in Tastle and Wierman (2007). Distribution A has 100 observations; distribution K has 10 observations. Case C reflects a situation defined as 0.5 by Van der Eijk, while the other measures do not have such a clearly identified midpoint. See Appendix B for a full description of the distributions, including the y-axes, although the table provides a characterization of the distributions in the left-most column. Not shown are the scaled versions of distributions K and U for which (by definition) only the standard deviation-based measures vary (Appendix C).

researchers mean by concentration. For example, in a highly skewed distribution where most of the mass is near a single extreme, such as distribution T, individual observations at the opposite extreme may have an outsize effect on the standard deviation because of their great distance from the mean.

Case A shows ideal-typical dispersion, with all observations in the extreme categories furthest left or right. Values of concentration should increase as this dispersion is reduced in the progression toward cases B and C, where the mean position is held constant. We can see in Table 2 that this is the case for all measures considered, and we note that the values of the standard deviations decrease with increasing concentration: all measures have face validity. Similarly, all measures identify case J as having the maximum concentration (and zero standard deviation). However, the ostensible midpoint between concentration and dispersion varies depending on the measure, and only measure *A* has an interpretable midpoint at 0.5. This means that studies interested in classifying outcomes in absolute terms—a tendency toward concentration versus a tendency toward dispersion—find it difficult to do so.

According to most measures, case F with a clear peak is more concentrated than case E, although *L2* and $L2^2$ disagree and regard case E as more concentrated. Similarly, most measures indicate greater concentration in case G than in case F because the share in the middle category is larger. In contrast, the standard deviation suggests no difference in concentration between the case with a single peak and that with two separate peaks. This is an ideal-typical instance that we have not seen discussed previously, where the standard deviation fails face validity as a measure of concentration.

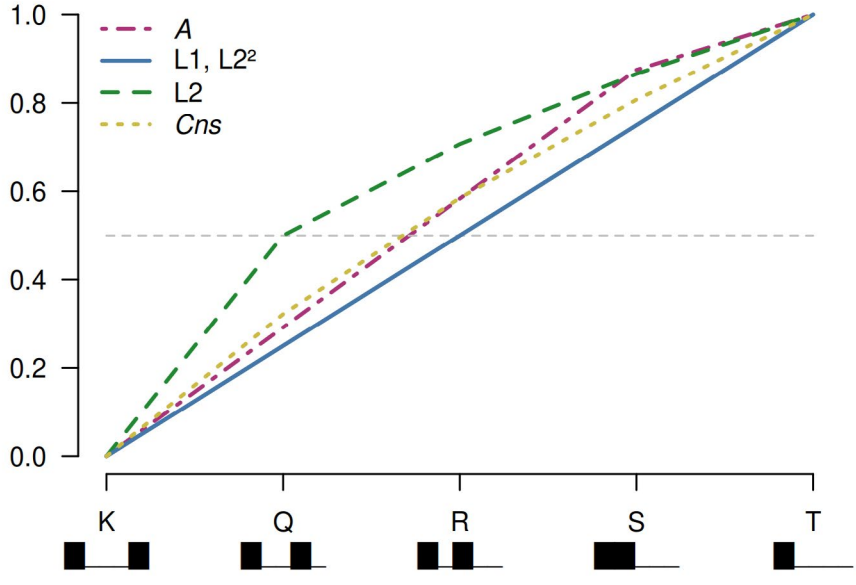
Cases H and I can be considered as equal concentrations, and all the measures considered yield the same result, including the standard deviation, further highlighting face validity and high correlations when considering ideal-typical cases. However, if we remove the assumption of equal spacing by taking the logarithm of the values, the results for the standard deviation vary. The log transformation can be thought of as adjusting for skewness in the distribution, which means that cases with the same standard deviation can have different log-transformed standard deviations if the underlying data distributions have different levels of asymmetry or skewness.

In Appendix C, we include additional cases used by Tastle and Wierman (2007), which demonstrate that all measures considered pass face validity. For instance, cases K to N in Appendix C are variations of case K in Table 2—half the observations in both extremes—but with an increasing number of observations; cases U to Y in Appendix C represent another series where the number of observations increases, but the distribution across categories remains the same. Only the standard deviation reacts to the change in the number of observations, which makes comparisons in concentration difficult across samples. The other measures do not share this undesirable property.

The different measures vary in the extent to which they react to minor differences in the data and differences in the order of magnitude of measurement error in many applications in the social sciences. These different reactions come to the fore in the series of cases {K, Q, R, S, T} in Figure 1. The series illustrates a change in distributions from dispersed in case K to concentrated in case T. All the measures reflect this change, providing face validity, but the size of the steps between cases varies. *L2* reacts most quickly to changes away from maximum dispersion (case K on the left of the figure). *Cns* and *A* react faster than the others (*L1*, $L2^2$), where changes seem linear. We note that changes away from the maximum concentration (case T on the right of the figure) are not symmetrical. A different test of how sensitive the different measures are to small changes in the data is provided by Case O in Appendix C: It changes 1% of the observations from one extreme category to the other. For 3 of the measures (*A*, *L1*, *L2*) and the standard deviation, this change is reflected in a change of number at the three-digit level, further highlighting that the different measures react differently to small changes.

Figure 1

Changes in Concentration in Simulated Response Distributions



Note. Concentration is given on the y-axis. There are 5 observations for each measure (x-axis), which are presented as lines to better visualize the differences.

The distributions in Table 2 are all based on 5 categories. This is an important detail to keep in mind. It follows from the definitions of each measure that split-

ting one of the categories into several categories generally changes the measured concentration. To illustrate this point, if we take distribution B and divide the empty middle category into three empty categories (see case B7 in Appendix D for concentration scores for 7 categories), all measures give numbers that differ from case B. Given that we are looking at ordinal data, it can easily be argued that dividing a category into subcategories should make no difference because case B7 may just (arbitrarily) make finer distinctions. Similarly, we note that dividing the flat distribution in case C into 7 categories affects all measures except for measure *A*, which is designed to have a fixed and interpretable midpoint (see Appendix D for concentration scores with 7 categories). To reiterate, scores cannot be compared when the number of categories varies, which includes recoding ordinal data into a smaller number of categories to simplify data or address small counts in some categories.

Correlations in these ideal-typical cases are remarkably high: $r \geq 0.93$ [0.85, 0.97] for all combinations that do not include the standard deviation divided by the number of observations (SD/n) or the logarithm (see Appendix E for a table of all correlations). The absolute value of the correlations with the standard deviation is high for all measures: 0.96 [0.92, 0.98] for *A*, 0.97 [0.93, 0.99] for *L1*, 0.96 [0.92, 0.98] for *L2*, 0.99 [0.97, 1.00] for $L2^2$, and 0.98 [0.96, 0.99] for *Cns*, even though some of them were designed to illustrate the weaknesses of the standard deviation for measuring concentration and dispersion in ordinal data.

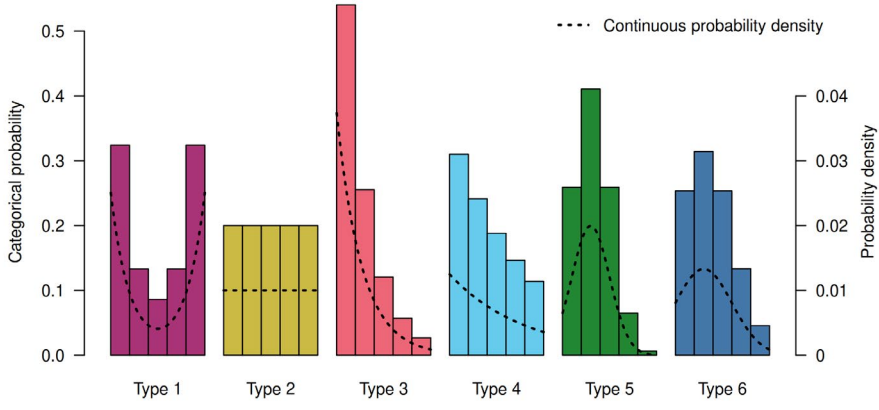
Simulations Assuming Continuous Latent Variables

In the second set of simulations, we assume a continuous latent variable that is divided into categories (see Agresti, 2002; Powers & Xie, 2008 for a discussion on understanding ordinal data as continuous). This provides us with a reference point that was not available in the preceding section, where no latent variable was assumed. We simulate responses by drawing 1,000 samples of size 500 from each of the four continuous distributions on the interval [0, 100]. We simulate 6 distinct types of distribution, as illustrated in Figure 2. We have chosen these types of distributions because they represent distributions often found in social science literature, particularly in situations where measures of distribution or concentration are of interest. Type 1 is a bimodal distribution defined as the sum of an exponential distribution with rate parameter $\lambda = 0.05$ and the reflection of this distribution about the point $x = 50$. Type 2 is the uniform distribution defined on [0, 100]. Type 3 is an exponential distribution with rate parameter $\lambda = 0.0375$, representing a unimodal distribution with small variance, whereas Type 4 is an exponential distribution with rate parameter $\lambda = 0.0125$, representing a unimodal distribution with larger variance. Type 5 is a unimodal distribution with small variance, defined by a normal distribution with a mean of 30 and

standard deviation 20, while Type 6 is a unimodal distribution with larger variance, defined by a normal distribution with a mean of 30 and standard deviation 30.

Figure 2

Simulated Response Distributions



Note. The dashed line gives the continuous probability density for 6 different simulated response distributions (Type 1, Type 2, …, Type 6); the bars break the probability density into a five-level ordinal variable.

Once the continuous distribution is simulated, we break the data into five-level ordinal variables with equal spacing. This corresponds to the commonly used Likert scales or survey rating scales, written {1, 2, 3, 4, 5}: a discretized version of the underlying continuous variable as it is often used in the social sciences to capture agreement at 5 distinct positions from 'strongly agree' to 'strongly disagree'. All points in the interval [0, 20] correspond to response 1, those in the interval [20, 40] correspond to response 2, and so on. In this sense, category frequencies are qualitative reflections of the continuous latent variable. After breaking the sample into these five bins, we calculate Leik's *D*, Blair and Lacy's *l*, Tastle and Wierman's *Cns*, Van der Eijk's *A*, and the categorical standard deviation. The choice of five bins allows a direct comparison with the examples in Van der Eijk (2001). We then compare these measures with the standard deviation calculated over the continuous sample.

The continuous latent variable model has two principal advantages. First, by assuming the existence of a continuous variable, we can model individual responses using well-defined continuous distributions. Second, there is already a well-respected measure of dispersion for continuous variables: the standard deviation. If we treat the continuous variable as a more truthful description of

respondents' positions, then the continuous standard deviation is an accurate summary of concentration (or dispersion) in the sample.[3] In the following analysis, ordinal measures that agree more closely with the continuous standard deviation are considered to be more valid.

Because the width of one category, say 0 to 20 for category 1, is not always substantively equivalent to the width of another category, we also conduct a logarithmic transformation on the simulated points,[4] which treats lower points as substantively further apart (see also Blair & Lacy, 2000). We then apply the different measures and the categorical standard deviation to the transformed continuous points.

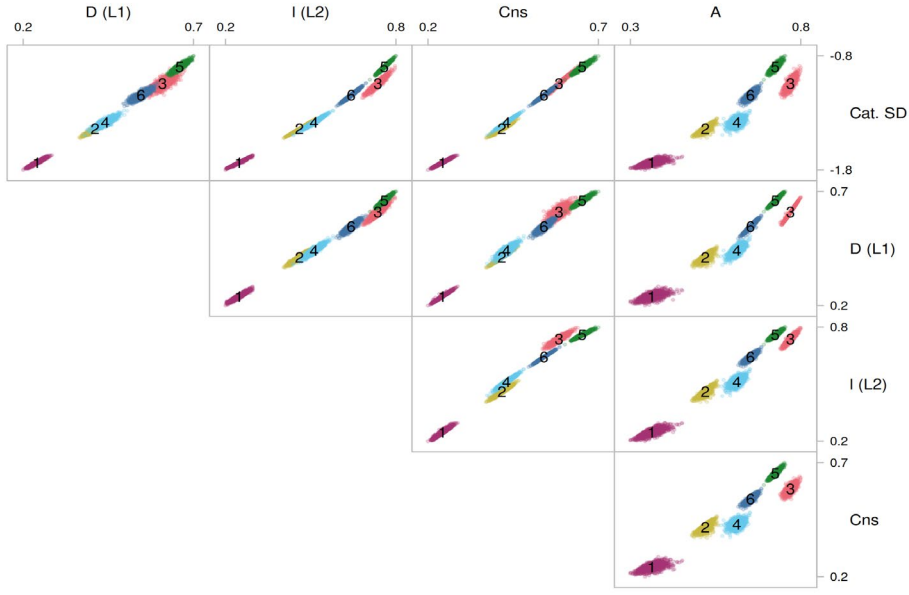
Initially, we draw on ideal-typical cases to offer examples of the extent to which different indices agree in their measurement of concentration. We consider the six ideal-typical distributions outlined in Figure 2 above. Figure 3 graphically shows the different values obtained according to the different measures and how they relate to one another. We can observe high correlations in these ideal-typical cases, but even though the four measures of concentration are standardized to the same range [0, 1], the scores vary widely by measure, as visible by their placement on the x-axis in Figure 3. These differences are so large that we adjusted the visible range on both axes for better readability, as indicated in the margins of the figure. For example, for the wide unimodal distribution (Type 4), the indices take values of $D = 0.571$, $\mathrm{l} = 0.697$, and $A = 0.690$. They vary even more for the tight unimodal distribution (Type 3), for which $Cns = 0.398$, $l = 0.479$, and $A = 0.683$. Van der Eijk designed A so that it equals 0.5 for the uniform distribution (Type 2), while the other indices vary between $D = 0.429$ and $l = 0.488$. Appendix F shows the correlations of the different measures with the linear standard deviations based on the latent continuous variables. Most correlations are above 0.85, with somewhat lower correlations for *L1* and notably *A*, which has lower correlations, especially for distributions of Type 1 (bimodal) and Type 4 (wide unimodal).

[3] We recognize that there are other measures of statistical dispersion, but in the following, we focus on the most widely used measure. In other cases, such as measuring inequality—rather than dispersion—in an outcome, measures such as the Gini index may be better suited.

[4] An alternative approach would be to cut the continuous distribution randomly, but we do not expect to gain additional insight from this computationally intensive approach.

Figure 3

Comparison of Measures in Simulated Cases



Note. The numbers in each panel indicate the ideal type, as outlined in Figure 3: 1 bimodal (purple), 2 uniform (yellow), 3 highly skewed (red), 4 skewed (cyan), 5 normal with peak (green), and 6 normal but flatter (blue). The scale differs for better readability. The colored areas behind the numbers show the distributions from 1,000 bootstraps as an indication of uncertainty. The four measures, *D*, *I*, *Cns*, and *A*, are standardized to [0, 1]. The top-left panel indicates the concentration level according to measure D for the six different ideal-typical distributions.

Figure 3 demonstrates that it is difficult to determine the concentration of a single distribution based on one of the approaches alone. To a substantial extent, this reflects the fact that, except for *A* at 0.5, the different measures provide little intuitive meaning beyond the endpoints 0 and 1. This problem is further complicated by the fact that some authors, such as Blair and Lacy (2000), prefer square indices. While squaring an index will result in a measure that is more sensitive to change when the measure is greater than 0.5, it does not help in understanding a single distribution. In the bimodal distribution (Type 1), for example, $I^2 = 0.051$, which ostensibly suggests that the distribution is highly dispersed. The other indices are all around 0.2, suggesting less dramatic dispersion. It follows that caution should be exercised when employing an index to describe the concentration of a single distribution in absolute terms.

Subsequently, we focus on the relative performance of the measures across different distributions of responses. In Figure 3, we can see that the spacing between the colored areas is inconsistent. Moreover, while most measures pro-

vide the same order (1, 2, 4, 6, 3, 5 starting at the bottom in each panel, focusing on differences on the x-axis), *A* yields a different order among the last three items (6, 5, 3 instead of 6, 3, 5). Measure *A* also differs notably from the others in that it clearly distinguishes between flat distribution (Type 2) and mild concentration (Type 4), something probably attributable to the algorithm's use of layers. We imagine that if there are preferences for these properties, they depend on the research question at hand. For the more concentrated distributions (Types 3, 5, and 6), we note different sensitivities: *L2* considers Types 3 and 5 more similar, *Cns* identifies more similarity between Types 5 and 6, while the standard deviation and *L1* space them equally. We expect that in real-world data, it will be difficult to judge the difference or similarity in concentration from the scores alone.

Empirical Example: Dispersion of Political Claims as Polarization

To better understand what the simulations mean in practice because simulated data may not fully capture the complexities of real-world data, we apply the different measures to data presented by Van der Brug et al. (2015) describing the position of actors toward immigrants in Western European newspaper articles. This introduces data such as those used by other researchers in their own studies and allows us to test whether the results from the ideal-typical cases have substantial implications in research.[5] We calculate the concentration of sentiment by year and country, allowing us to observe trends in concentration between 1995 and 2009. Although data are given for seven European countries, we compare two in the main text: Austria, in which the indices correlate closely, and the United Kingdom, in which the indices imply different longitudinal trends.

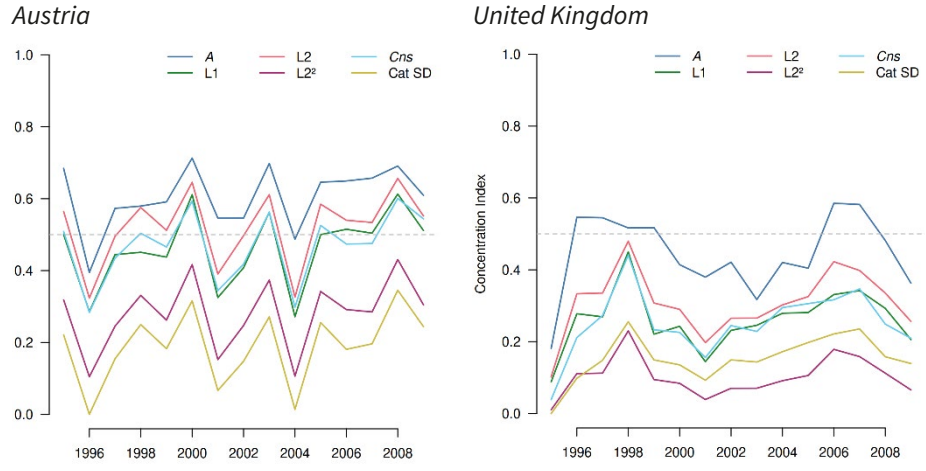
Van der Brug et al. (2015) coded claims about immigration in newspapers and assigned a positional score to each claim to identify whether the claim was positive or negative, using 5 categories. They then calculated Van der Eijk's measurement *A* for claims in a country and year to trace dispersion over time as a measure of polarization.[6] In Figure 4, we reproduce two of these figures with different measures (see Appendix G for the other 5 countries, which constitute intermediate cases).

[5] While the articles are randomly sampled, it is debatable to what extent the coding of positions is a probability sample. However, we focus on the concrete implications of using different measures for researchers who may willingly violate this assumption.

[6] We refer readers to the literature on measuring polarization (e.g., DiMaggio et al. 1996) to decide whether dispersion as captured by *A* is adequate to capture polarization.

Figure 4

Comparison of Concentration in Austria and the United Kingdom



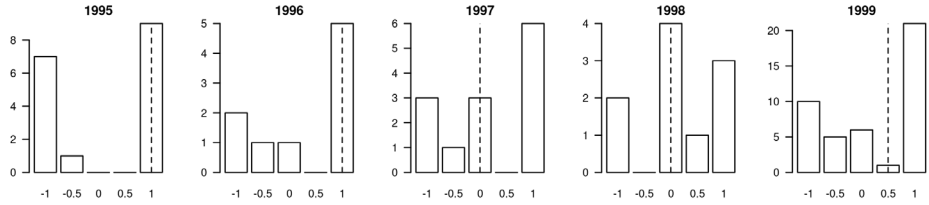
Note. Concentration of position of claims in national newspapers for Austria (left panel) and the United Kingdom (right panel), 1995 to 2009, measured using 5 response categories and different measures of concentration. Using Van der Eijk's measure A, the dashed line at 0.5 indicates neither concentration nor dispersion.

We note two aspects. First, Van der Brug et al. (2015) interpreted the values from Van der Eijk's measure *A* using the dashed line at 0.5 as a reference indicating neither concentration nor dispersion, or in their terms, polarization and agreement. This comparison, supported by Van der Eijk (2001), allows them to identify tendencies toward concentration and dispersion in absolute terms. The placement of the lines in Figure 4 makes it clear that such an interpretation is not possible with the other measures. More generally, looking at the height of the lines in the figure gives us little sense of whether in a specific year media reporting was particularly concentrated or dispersed.

Second, in Austria, given on the left of Figure 4, the different indices correlate closely, while in the United Kingdom, given on the right of the figure, the indices imply different longitudinal trends. For instance, while measure *A* suggests no substantive change in concentration between 1996 and 1999 (blue line at the top), the other indices suggest a peak in concentration in 1998. Inspecting the graph, we can identify a general decrease between 1997 and 2003 using *A* but a decrease within the notably shorter 1998 to 2001 period. If these values are related to events such as elections or refugee flows, substantially different conclusions can ensue.

Figure 5

Concentration in the UK, 1995 to 1999



Note. Positions on immigration are measured in 5 categories: -1, -0.5, 0, +0.5, +1. Horizontal dashed lines indicate the median position.

Figure 5 shows the distribution of the positions of claims in the UK for each year between 1995 and 1999 to understand why the different measures of concentration disagree on the changes over time. The situation in 1995 corresponds to a Type 1 distribution with clear dispersion (close to case *A* in Table 1). In 1996 and 1997, there was a concentration of claims with position +1, corresponding to Type 3. In this case, the categorical standard deviation, *A*, and *L1* and *L2* scores suggest a slight difference in concentration. In contrast, the measure *Cns* identifies a marked difference, with more concentration in 1997. In 1998, there was a concentration at position 0, with secondary peaks of roughly equal size at +1 and -1. In this case, *A* differs from the other measures. Compared with 1997, these two measures identify a slight decrease in concentration, while the others suggest a slight increase. In 1999, the distribution of position was back at Type 3 with a concentration at position +1. Yet, despite the visual similarity to 1996, measures *L1* and *L2* clearly differ. We suspect that the way different measures of concentration respond to subtle differences in distributions may not be obvious and predictable to social scientists.

Discussion and Conclusion

Ordinal data are common in the social sciences—just think of the Likert items in surveys, rating scales, or content analysis—and so are questions about the concentration and dispersion of these data. Researchers across several fields have proposed measures of concentration and dispersion with ordinal data; however, there has been little work to compare these methods.

Here we compare the different approaches to better understand their properties and how they behave in practice. Overall, we found high correlations between the measures, especially when considering ideal-typical cases. This implies that in many situations, any measure can be used. However, minor dif-

ferences in distributions can yield distinct conclusions. For instance, the conclusions for the 1998 UK politicization data differ across measures, and in the simulations, whether we conclude that an exponential distribution (Type 3) is more concentrated than a unimodal distribution (Type 5) depends on the measure used. Because some of these differences depend on where we are on the scale between concentration and dispersion, we suspect that they are difficult for researchers to anticipate.

Starting with the high correlations we observed, this includes the standard deviation dismissed by the authors of alternative measures for both constructed and real research data. This remains the case when we use the logarithm to vary the spacing between categories, undermining a key argument in favor of alternative measures. In this sense, researchers who gloss over the dangers of the standard deviation when applied to ordinal data may not necessarily draw incorrect conclusions about the concentration of their data: typically, lower standard deviations correspond to higher concentrations. More generally, the different measures we considered tend to agree which distributions are more concentrated, but they do not agree on the extent of concentration. In other words, if we consider a change from a concentrated to a dispersed distribution, the different measures agree that change occurred (because they are all consistent in that they react to shifting 'mass' toward the poles), but they do not agree about the size of the steps. As we have shown, the different measures typically lead to the same rankings—although measure *A* can differ—but not the same spacing nor the same scores outside the endpoints.

Even though (apart from the standard deviation) the measures use the same scale from 0 to 1, we warn researchers against interpreting the scores in absolute terms. Because of the different spacing and because they react differently to minor changes in the distributions, the absolute positions have no intuitive meaning beyond the endpoints (*A* is an exception with an interpretable midpoint at 0.5). Instead, the different measures should be reserved for the comparison of distributions in relative terms: they are reliable in identifying which of two distributions is more concentrated or dispersed.

Scores from different measures cannot be compared, nor can we compare scores from 5-point and 7-point scales. In practice, this means that recoding ordinal data (for example, if there are few observations in some of the categories) can have a direct impact on the conclusions we draw in terms of concentration and dispersion. If categories are recoded, we recommend comparing concentration and dispersion before and after recoding and discussing differences if they occur.

Depending on the conceptualization or understanding of concentration and dispersion in a particular project, researchers may want to prioritize one of the measures (Appendix H). Where prioritization is not apparent, but also more generally as tests of robustness, we recommend comparing multiple measures and

discussing cases where they differ from one another in qualitative terms. This is necessary because they react differently to changes in distributions; therefore, the measures disagree whether differences between distributions are substantially important or statistically significant for researchers concerned with p-values. Put differently, if the conclusions we draw depend on the measure used, we have less confidence in the results. We see no obvious reason to exclude the standard deviation from the suite of indices used to analyze differences in ordinal concentration.

Based on our analysis, we provide the following recommendations:

- All measures of concentration and dispersion considered are suitable for describing changes in distributions, including the standard deviation with which most readers are probably familiar.
- Concentration and dispersion scores have little intuitive meaning beyond the ideal-typical endpoints (and the midpoint in the case of measure *A*).
- Interpretation in absolute terms should be avoided, and the measures should instead be used to compare sets of responses on the same scale. Despite the high correlations between scores, we note that differences occur in specific cases; therefore, we recommend using multiple measures of concentration—coupled with visual inspection of distributions where the measures disagree.
- Scores from different measures should not be directly compared. They are not interchangeable.
- Scores from a distribution with a different number of categories (e.g., 5-point versus 7-point scale) should not be compared but can be used to test the robustness of substantive findings.

When considering different measures, we consider it desirable to have hard boundaries—something where the standard deviation falls short—but we have noted that only the ideal-typical endpoints have consistent values across the measures, in which case we do not really need numerical summaries. For the intermediate situations, however, the different measures make it difficult to tell how close we are to the ideal points of concentration and dispersion, despite all being able to identify changes in the distribution. A visual examination of the distributions can help understand the behavior of the indices.[7] This is important because we suggest that the importance of different features may depend on the topic under study.[8]

---

[7] A visual examination can also differentiate cases of dispersion that the measures consider the same because they capture dispersion without regard to the position: a distribution of (5, 0, 5, 0, 0) is considered indistinguishable from (0, 5, 0, 5, 0) in terms of dispersion, even though when using dispersion to capture polarization, researchers may consider the latter as more polarized.

[8] With a larger number of categories, differences in continuous variables may become less pronounced, and researchers may want to consider the difference between skew and kurtosis when deciding how to measure dispersion.

To conclude, no single measure of concentration and dispersion can serve all cases, and we must be careful with interpretations. In this sense, it is important to understand the limits of the different measures and how they react to different changes in distributions. For broad changes all the measures we considered should provide equivalent results. However, researchers may have reasons to prioritize one of the measures for how they react to changes in concentration. Outside such preferences, we recommend using a range of measures of concentration or dispersion to ascertain the robustness of substantive findings and visual assessments of the distributions, especially in cases where measures disagree in the extent of difference.

Data and Code Availability

An R package to calculate concentration and dispersion is available (https://doi.org/10.32614/CRAN.package.agrmt), as is replication material (https://doi.org/10.17605/OSF.IO/UMBHD).

Conflict of Interest

The authors declare no conflict of interest.

Ethics

No human subjects were involved. Exempt according to institutional guidelines.

Funding

This work was supported by the Brown University LINK Internship Award [to CA]; the NCCR on the move [grant numbers 182897 and 205605]; and the Swiss National Science Foundation [grant number 200939].

Contributions

Authors are listed alphabetically. DR designed the study, CA and DR carried out the simulations, CA and DR carried out the analysis, DR and CA wrote the paper.

References

Agresti, A. (2002). *Categorical data analysis (Vol. 792)*. John Wiley & Sons. https://doi.org/10.1002/0471249688

Alwin, D. F., & Tufiş, P. A. (2016). The changing dynamics of class and culture in American politics: A test of the polarization hypothesis. *The ANNALS of the American Academy of Political and Social Science, 663*(1), 229–269. https://doi.org/10.1177/0002716215596974

Anderson, L. R. (2025). An unreliable ladder: Top-bottom self-placement, subjective social status, and political preferences. *Sociological Science, 12*(25), 601–633. https://doi.org/10.15195/v12.a25

Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. *Proceedings of the National Academy of Sciences, 118*(52), e2112552118. https://doi.org/10.1073/pnas.2112552118

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. CRC Press.

Berg, M. T., Stewart, E. A., Stewart, E., & Simons, R. L. (2013). A multilevel examination of neighborhood social processes and college enrollment. *Social Problems, 60*(4), 513–534. https://doi.org/10.1525/sp.2013.60.4.513

Berry, K., & Mielke, P. (1992). Assessment of variation in ordinal data. *Perceptual and Motor Skills, 74*(1), 63–66. https://doi.org/10.2466/pms.1992.74.1.63

Blair, J., & Lacy, M. (2000). Statistics of ordinal variation. *Sociological Methods & Research, 28*(3), 251–280. https://doi.org/10.1177/0049124100028003001

Davies, V. (1970). A re-examination of Leik's measure of ordinal consensus. *The Pacific Sociological Review, 13*(1), 27–28. https://doi.org/10.2307/1388402

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology, 102*(3), 690–755. https://doi.org/10.1086/230995

Enders, A. M. (2021). Issues versus affect: How do elite and mass polarization compare? *The Journal of Politics, 83*(4), 1872–1877. https://doi.org/10.1086/715059

Haas, S., & Hüllermeier, E. (2025). Uncertainty quantification in ordinal classification: A comparison of measures. *International Journal of Approximate Reasoning, 186*, 109479. https://doi.org/10.1016/j.ijar.2025.109479

Harding, D. J. (2011). Rethinking the cultural context of schooling decisions in disadvantaged neighborhoods: From deviant subculture to cultural heterogeneity. *Sociology of Education, 84*(4), 322–339. https://doi.org/10.1177/0038040711417008

Krymkowski, D. H., Manning, R. E., & Valliere, W. A. (2009). Norm crystallization: Measurement and comparative analysis. *Leisure Sciences, 31*(5), 403–416. https://doi.org/10.1080/01490400903199443

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association, 45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Kvålseth, T. O. (1988). Measuring variation for nominal data. *Bulletin of the Psychonomic Society, 26*(5), 433–436. https://doi.org/10.3758/BF03334906

Kvålseth, T. O. (1995). Coefficients of variation for nominal and ordinal categorical data. *Perceptual and Motor Skills, 80*(3), 843–847. https://doi.org/10.2466/pms.1995.80.3.843

Leik, R. (1966). A measure of ordinal consensus. *Pacific Sociological Review, 9*(2), 85–90.

Levendusky, M. S. (2009). The microfoundations of mass polarization. *Political Analysis, 17*(2), 162–176. https://doi.org/10.1093/pan/mpp003

Lupu, N. (2015). Party polarization and mass partisanship: A comparative perspective. *Political Behavior, 37*(2), 331–356. https://doi.org/10.1007/s11109-014-9279-z

Powers, D., & Xie, Y. (2008). *Statistical methods for categorical data analysis.* Emerald Group Publishing.

Reardon, S. (2009). Measures of ordinal segregation. *Research on Economic Inequality, 17,* 129–155. https://doi.org/10.1108/S1049-2585(2009)0000017011

Tastle, W., & Wierman, M. (2007). Consensus and dissention: A measure of ordinal dispersion. *International Journal of Approximate Reasoning, 45*(3), 531–545. https://doi.org/10.1016/j.ijar.2006.06.024

Tran, D., Dolgun, A., & Demirhan, H. (2020). Weighted inter-rater agreement measures for ordinal outcomes. *Communications in Statistics – Simulation and Computation, 49*(4), 989–1003. https://doi.org/10.1080/03610918.2018.1490428

Van der Brug, W., D'Amato, G., Berkhout, J., & Ruedin, D. (Eds.). (2015). *The Politicisation of Migration.* Routledge. https://doi.org/10.4324/9781315723303

Van der Eijk, C. (2001). Measuring agreement in ordered rating scales. *Quality and Quantity, 35*(3), 325–341. https://doi.org/10.1023/A:1010374114305

Appendix A

Calculating

The following describes how the agreement score is calculated by Van der Eijk (2001). The frequency vector represents a 7-point ordinal variable: (30, 40, 210, 130, 530, 50, 10). This is divided into layers, where each response category has the same number of observations. The first layer has the size of the smallest category ($n = 10$), and no empty categories (pattern 1111111). The number of non-empty categories in the pattern is $S = 7$. The weight of the layer is

$$\frac{L = 7 \text{ categories} * n = 10 \text{ observations}}{R = 1000 \text{ responses} \in \text{vector}} = 0.07 .$$

The agreement of each layer is calculated as

$$A_{\text{layer}} = 1 - \frac{S\text{-}1}{L\text{-}1},$$

where S is the number of non-empty categories in a layer, and L is the total number of response categories ($L = 7$ in this example). For layer 2, the number of non-empty categories in the pattern is $S = 6$. The agreement of the layer is thus 0.0167. The agreement score A for the entire vector is the weighted average of the layer agreements. We refer the reader to the original paper for a more thorough explanation of the algorithm.

Table A1

Van der Eijk's Algorithm Divides the Vector into Layers, and Identifies Non-Empty Categories in Each Layer as the Basis of a Weighted Agreement Score

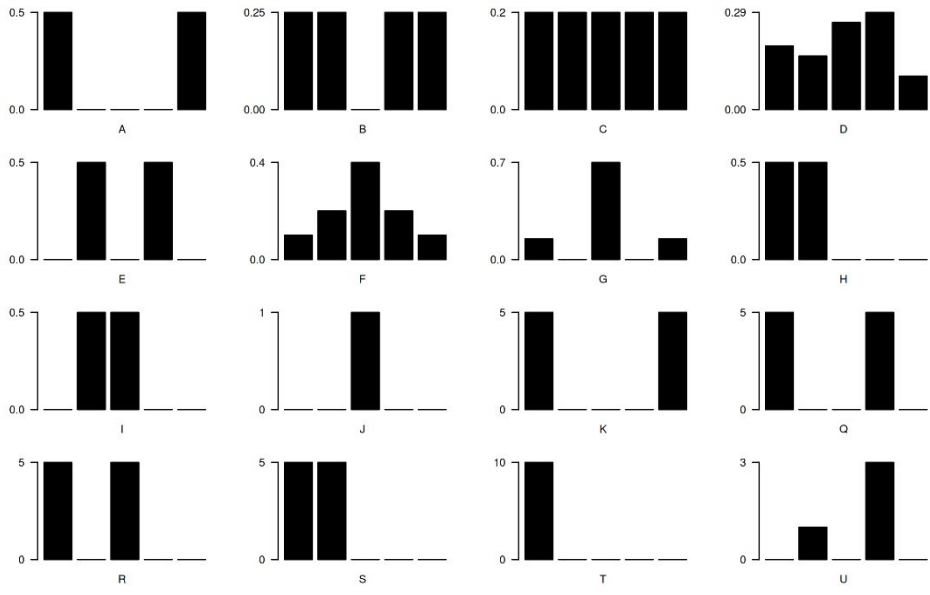
Layer	1	2	3	4	5	6	7	Pattern	Layer agreement	Weight
Layer 1	10	10	10	10	10	10	10	1111111	0	0.07
Layer 2	20	20	20	20	20	20	0	1111110	0.167	0.12
Layer 3	0	10	10	10	10	10	0	0111110	0.333	0.05
Layer 4	0	0	10	10	80	10	0	0011110	0.5	0.04
Layer 5	0	0	80	80	80	0	0	0011100	0.667	0.24
Layer 6	0	0	80	0	80	0	0	0010100	0.467	0.16
Layer 7	0	0	0	0	320	0	0	0000100	1	0.32
Total	30	40	210	130	530	50	10			

Appendix B

Details on the Distributions

Figure B1

Ideal-Typical Distributions from Tastle and Wierman



Note. Graphical presentation of the ideal-typical distributions used by Tastle and Wierman (2007), which we labeled consecutively as A to Y. Cases A – J are from Table 1 in Tastle and Wierman. Their Table 2 includes scaled versions of case K included in this figure (cases K to N in Appendix C, distribution O in the supplement differs from K by only 1%); their Table 3 corresponds to the series {K, Q, R, S, T} (distribution P is equal to distribution K). Cases U to Y from Table 4 in Tastle and Wierman are scaled versions of the distribution U included in this figure (compare Appendix C). Case D does not represent an ideal-typical distribution, but is one of their examples.

Appendix C

Additional Distributions

Distributions from Tastle and Wierman that are scaled versions of each other or differ only marginally (distribution O) are not included in Table 2. Table C2 shows that these (by definition) only vary in the measures derived from the standard deviation.

Table C2

Concentration Scores for the Rescaled Distributions

Distribution	*A*	*L1*	*L2*	$L2^2$	*Cns*	*SD*	ln(*SD*)
K ▄_▁_▄	0.000	0.000	0.000	0.000	0.000	2.108	0.848
L ▄_▁_▄	0.000	0.000	0.000	0.000	0.000	2.010	0.809
M ▄_▁_▄	0.000	0.000	0.000	0.000	0.000	2.001	0.805
N ▄_▁_▄	0.000	0.000	0.000	0.000	0.000	2.000	0.805
O ▄_▁_▄	0.020	0.020	0.020	0.000	0.000	2.010	0.809
U _▄_▄_	0.792	0.750	0.791	0.625	0.686	1.000	0.347
V _▄_▄_	0.792	0.750	0.791	0.625	0.686	0.877	0.304
W _▄_▄_	0.792	0.750	0.791	0.625	0.686	0.871	0.302
X _▄_▄_	0.792	0.750	0.791	0.625	0.686	0.870	0.301
Y _▄_▄_	0.792	0.750	0.791	0.625	0.686	0.866	0.300

Note. Rescaled distributions that are not included in Figure 3. The highest bar in distribution K is 5 (the histogram is scaled differently from Table 2), in distribution L 50, in distribution M 500, in distribution N 5,000. Distribution O resembles distribution L, but one bar is 2 larger than the other. For distribution U, the non-zero bars are 1 and 3, for distribution V 10 and 30, for distribution W 20 and 60, for distribution X 30 and 90, and for distribution Y 300 and 900.

Appendix D

Seven Categories

Change from maximum dispersion to maximum concentration for 5 categories, and for 7 categories. The cases correspond to those presented in Figure B1, with the number 7 highlighting the use of 7 categories. Where the use of 7 categories required a distinction, these are indicated with letters 'a' and 'b'.

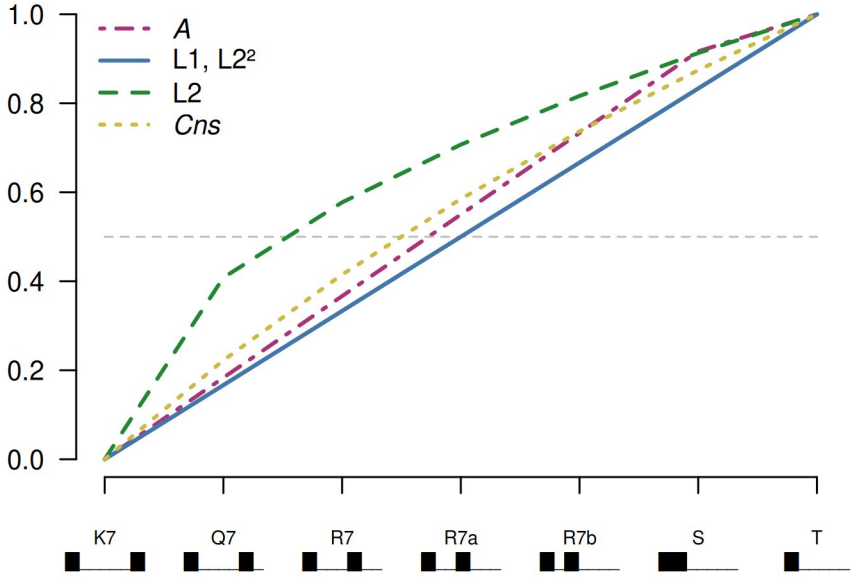
Table D3

Concentration Scores for Distributions Based on 7 Categories

Distribution	*A*	*L1*	*L2*	$L2^2$	*Cns*
K7 ▮_____▮	0.000	0.000	0.000	0.000	0.000
Q7 ▮____▮_	0.183	0.167	0.408	0.167	0.222
R7 ▮___▮__	0.367	0.333	0.577	0.333	0.415
R7a ▮__▮___	0.550	0.500	0.707	0.500	0.585
R7b ▮_▮____	0.733	0.667	0.816	0.667	0.737
S7 ▮▮_____	0.917	0.833	0.913	0.833	0.874
T7 ▮_____	1.000	1.000	1.000	1.000	1.000
A7 ▮_____▮	0.000	0.000	0.000	0.000	0.000
B7 ▮▮___▮▮	0.383	0.167	0.289	0.083	0.208
B7b ▮▮▮_▮▮▮	0.473	0.333	0.430	0.185	0.384

Figure D2

Changes in Concentration for 7 Categories



Note. Compare Figure 1 for 5 categories.

Appendix E

Correlations

Table E4

Correlation between Measures of Concentration

Measure	*A*	*L1*	*L2*	$L2^2$	*Cns*	*SD*	ln(*SD*)
A	1.00	0.99	0.99	0.96	0.99	−0.96	−0.95
L1	0.99	1.00	0.98	0.97	0.99	−0.97	−0.97
L2	0.99	0.98	1.00	0.96	0.99	−0.96	−0.94
$L2^2$	0.96	0.97	0.96	1.00	0.97	−0.99	−0.97
Cns	0.99	0.99	0.99	0.97	1.00	−0.98	−0.96
SD	−0.96	−0.97	−0.96	−0.99	−0.98	1.00	0.97
SD (log)	−0.95	−0.97	−0.94	−0.97	−0.96	0.97	1.00
SD/n	−0.26	−0.25	−0.21	−0.22	−0.26	0.31	0.33

Note. Based on cases A to Y, Pearson correlation coefficients.

Appendix F

Correlations with the Linear Standard Deviation

Table F5

Correlation with the Linear Standard Deviation by Measure and Distribution Type

Measure	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
Categorical *SD*	−0.855	−0.902	−0.953	−0.917	−0.877	−0.911
L1	−0.822	−0.865	−0.753	−0.800	−0.774	−0.776
L2	−0.847	−0.901	−0.911	−0.900	−0.862	−0.900
Cns	−0.842	−0.881	−0.943	−0.911	−0.826	−0.903
A	−0.585	−0.707	−0.749	−0.443	−0.753	−0.727

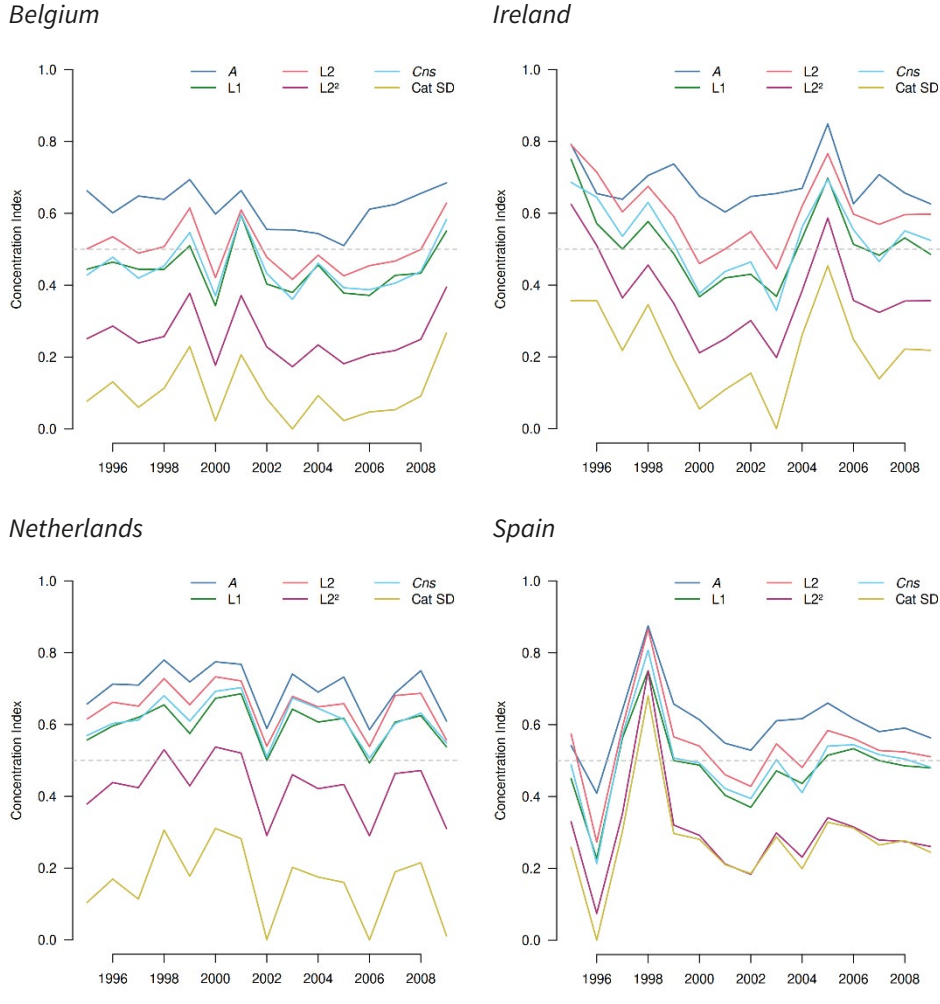
Note. Pearson correlations with the linear standard deviations (simulations as in Figure 3), see Figure 2 for distributions and latent continuous distributions.

Appendix G

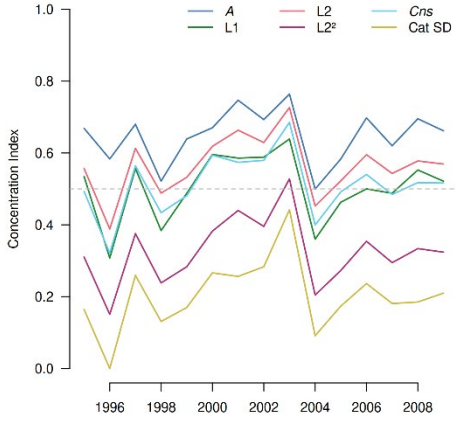
Country Distributions

Figure G3

Polarization for Belgium, Ireland, Netherlands, Spain, Switzerland, 1995 to 2009



Switzerland



Note. Refer to Figure 4 for a description

Appendix H

Prioritizing Measures

If we want to detect small changes away from maximum dispersion, *L2* is preferable, but this feature does not apply symmetrically for small changes away from maximum concentration. If a change from a flat distribution to a mild concentration is important, measure *A* emphasizes these more than the others. If we want to de-emphasize differences from extreme peaks, e.g., because of anchoring effects, *L2* should be prioritized over other measures. Three measures (*L1*, *A*, *Cns*) emphasize the central category for concentration, while *L2* and $L2^2$ do not. Measure *A* is distinctive with a midpoint that is interpretable in absolute terms; however, its definition of the midpoint as the uniform distribution may not align with the way researchers understand concentration and dispersion in all applications. Finally, if the distinction between exponential and unimodal distributions is a central part of the understanding of concentration and dispersion, this is better captured by measures other than *L2*. In contrast, *Cns* is less suited for distinguishing between unimodal distributions with different variance.

Table H1

Considerations when Prioritizing Properties for Concentration Measures for Ordinal Data

Property	Prioritize	Avoid
Interpretable midpoint	*A*	others
Same score when sample size changes	others	*SD*
Hard boundaries (0, 1)	others	*SD*
Emphasize small changes	*L2*	
Emphasize differences between flat and mild concentration	*A*	*Cns*
De-emphasize difference from extreme peaks	*L2*	*A*
Emphasize central category for concentration	*A, L1, Cns*	*L2, L2*[2]
Distinguish unimodal from exponential distribution	*L1, Cns*	*L2*
Differentiate unimodal distributions with different variance		*Cns*

Note. None of the measures is suitable to capture changes in absolute terms, allows differences to be anticipated easily in all cases, or is safe for recoding (5 categories, 7 categories). Reactions to small changes are not symmetrical when measured with *L2*.