Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When the P Set Exceeds the Q Sample: Rank Deficiency, Eigenvalue Inflation and Purposive Sampling in Q Methodology

Raffaele Zanoli

zanoli@agrecon.univpm.it

Marche Polytechnic UniversityDe

Research Article

Keywords: Q methodology, P set, Q sample, Factor retention, Purposive sampling, Saturation, Rank deficiency

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11121183/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

A persistent question in Q methodology is whether the number of participants in the P set may exceed the number of items in the Q sample. Rules of thumb imported from R-technique factor analysis, which require more observations than variables, suggest that it may not. This note argues that the prohibition is unfounded, but that the opposite claim—that the ratio of persons to statements is analytically irrelevant—is equally untenable. Because Q factor analysis correlates persons across items, the person correlation matrix has rank at most Q − 1; whenever P ≥ Q it is necessarily singular, and its non-zero eigenvalues must absorb a total variance of P. Three consequences follow. First, the Kaiser eigenvalue-greater-than-one criterion becomes progressively less discriminating as P/Q grows, because the mean non-zero eigenvalue is at least P/(Q − 1); a simulation of structureless forced sorts shows that with Q = 40 and P = 200 the criterion retains all 39 extractable components from pure noise, and that where structure is present its count tracks the noise level rather than the number of latent viewpoints, whereas a parallel analysis built on permutations of the study's own forced distribution recovers that number. Second, procedures requiring the inverse of the person correlation matrix—maximum-likelihood extraction, Bartlett's test of sphericity, the Kaiser–Meyer–Olkin index and regression-based factor score estimation—become unavailable by construction, whereas principal components and centroid extraction are unaffected. Third, the standard error of an individual factor loading depends on Q alone, so additional participants do not sharpen individual loadings, although they can improve composite factor arrays when they supply clean defining sorts. Beyond these mathematical points, large P sets raise a design question, since additional participants may reproduce configurations already observed rather than extend the range of subjectivities encountered. P-set construction is therefore better governed by purposive coverage of theoretically relevant positions than by numerical representativeness or by a statistical stopping rule.

1 Introduction

A recurrent question in Q methodology concerns the appropriate relationship between the number of persons in the P set and the number of items in the Q sample. In conventional R-technique factor analysis considerable attention is given to the ratio between observations and variables, and rules of thumb commonly require substantially more observations than variables. When such reasoning is transferred to Q methodology it can lead to the assumption that a Q study becomes statistically problematic when the number of persons exceeds the number of statements.

That conclusion does not follow. Q factor analysis reverses the conventional orientation of factor analysis: persons, represented by their complete Q sorts, are correlated across the items of the Q sample (Stephenson 1953). The mathematical role of the P set is therefore fundamentally different from the role of sample size in R methodology. There is no general requirement that P < Q, and a study does not become invalid when P > Q (Brown 1980; Watts and Stenner 2012; McKeown and Thomas 2013).

The opposite conclusion, however—that the ratio of persons to statements is mathematically irrelevant—is also too strong. Once the number of Q sorts exceeds the number of statements, several properties of the person correlation matrix, of its eigenvalues and of the conventional criteria for factor retention change in predictable ways, and a number of standard multivariate procedures cease to be computable. Moreover, very large P sets raise a second, methodological issue: additional participants may increasingly reproduce configurations already present in the study rather than extend the range of subjectivities encountered. The issue is therefore not only how many persons can be factor analysed, but what additional persons contribute to the purposive logic of the P set.

Because this note is addressed to a general methodological readership rather than to Q specialists alone, it may help to fix terminology at the outset. The ‘concourse’ is the universe of statements that circulate about a topic; the ‘Q sample’ is the structured subset of Q items drawn from it; the ‘P set’ is the group of P participants recruited to sort those items; the ‘condition of instruction’ is the prompt under which they sort; a ‘Q sort’ is one participant's ranking of the Q items, normally into a forced quasi-normal distribution; a ‘defining sort’ is a Q sort that loads significantly on one factor and on no other; and a ‘factor array’ is the weighted composite Q sort that represents a factor. Throughout, P denotes the number of participants and Q the number of statements.

The contribution of this note is threefold. It states precisely, and in a form usable by applied researchers, the rank and eigenvalue consequences of the design P > Q; it identifies which analytical procedures those consequences rule out and which they leave untouched, a point that appears not to have been made explicitly in the Q literature; and it connects the resulting statistical picture to the purposive logic by which a P set is constructed, including the status of saturation as a stopping criterion. The argument is developed analytically and illustrated by a small simulation of structureless Q sorts.

2 The structure of Q factor analysis

2.1 Persons as variables, statements as observations

Suppose a study contains P persons sorting Q statements. Let X denote the resulting P × Q matrix, with rows representing persons and columns representing statements. Q factor analysis begins by correlating the rows of this matrix. The resulting person-by-person correlation matrix R therefore has dimension P × P.

The correlation between two persons is estimated from their respective scores across the Q statements. In this specific mathematical sense the statements provide the observations from which similarity between persons is estimated. Increasing P adds another Q sort to be compared with the existing Q sorts; it does not add another observation to the calculation of any existing person-by-person correlation. Increasing Q, by contrast, provides additional item observations from which those correlations are estimated.

2.2 Rank deficiency when P ≥ Q

Let Z denote the matrix obtained from X by standardising each row, that is, by centring each Q sort on its own mean and dividing by its own standard deviation. The person correlation matrix is then

R = (1 / (Q − 1)) ZZ'. (1)

Standardisation is worth making explicit, because it is sometimes ignored. Under a common forced distribution every Q sort has the same mean and the same variance, so centring alone already yields R up to a single scaling constant; under a free distribution the row-wise division by the standard deviation is required. The rank argument that follows is unaffected by scaling and therefore holds in both cases.

Because each Q sort has been centred, its Q scores sum to zero. The rows of Z therefore lie in a linear subspace of dimension no greater than Q − 1, and consequently

rank(R) = rank(ZZ') ≤ min(P, Q − 1), (2)

with equality holding generically, that is, for any set of Q sorts not lying in some lower-dimensional subspace by accident of design or duplication. Whenever P ≥ Q the P × P person correlation matrix must therefore be singular, and when P > Q there must be at least P − Q + 1 zero eigenvalues.

Consider a Q study involving 80 persons and 40 statements. The correlation matrix is an 80 × 80 matrix, but its rank is at most 39, so at least 41 of its eigenvalues must be exactly zero. This is not evidence that the study is underpowered or improperly specified. If the substantive objective is to identify three, four or five interpretable factors, a ceiling of 39 dimensions is immaterial. The singularity is simply a mathematical property of the design, and the number of factors that can be extracted is bounded by Q − 1 rather than by P.

The converse case is worth stating for completeness. When P < Q − 1 the person correlation matrix is generically of full rank P and non-singular, all P eigenvalues may be non-zero, and the accounting described in the next section does not apply. The issues raised here are specific to designs in which P approaches or exceeds Q.

3 Consequences for factor retention

3.1 Eigenvalue inflation and the Kaiser criterion

For a correlation matrix of P persons, trace(R) = P, and the eigenvalues therefore sum to P. When P > Q, however, no more than Q − 1 of them can be non-zero. The total variance represented by P must consequently be distributed across at most Q − 1 dimensions, so that the mean of the non-zero eigenvalues is at least

$\bar{\lambda} \geq P / (Q - 1)$. (3)

For P = 80 and Q = 40 this lower bound is 80/39 ≈ 2.05; for P = 200 and Q = 40 it is 200/39 ≈ 5.13. An eigenvalue exceeding 1 is therefore in no way remarkable in such a design.

This exposes a specific weakness in the mechanical application of the Kaiser eigenvalue-greater-than-one criterion (Kaiser 1960) to Q studies with large P/Q ratios. The rationale of that criterion is that, for a correlation matrix of n variables, the eigenvalues sum to n and their mean is therefore exactly 1, so that a component with an eigenvalue above 1 accounts for more variance than a single variable does. In a Q design with P > Q the mean over all P eigenvalues is still 1, but P − Q + 1 of them are structurally zero. Within the (Q − 1)-dimensional subspace in which factors can actually be extracted, the relevant mean is P/(Q − 1). The threshold of 1 therefore sits at a fraction (Q − 1)/P of the average extractable eigenvalue, and it becomes a progressively less demanding criterion as P grows while Q is held fixed. Table 1 makes the arithmetic explicit.

Table 1
Rank and eigenvalue consequences of the ratio of persons to statements

P	Q	Order of R	Max. rank	Min. zero eigenvalues	Mean non-zero eigenvalue	Loading threshold (.01)
20	40	20 × 20	20	0	1.00	0.408
40	40	40 × 40	39	1	1.03	0.408
80	40	80 × 80	39	41	2.05	0.408
100	40	100 × 100	39	61	2.56	0.408
200	40	200 × 200	39	161	5.13	0.408
100	60	100 × 100	59	41	1.69	0.333

The mean non-zero eigenvalue is the lower bound P/min(P, Q − 1) implied by trace(R) = P. The loading threshold is the conventional two-tailed .01 criterion 2.58/√Q. Note that the threshold depends on Q alone and is unaffected by P.

Eigenvalues still carry information. Their relative pattern, the shape of the scree and their comparison with an appropriate null distribution remain useful. What loses its meaning is the absolute threshold of unity.

3.2 A numerical demonstration

The practical force of the preceding argument can be shown directly. Q sorts were generated in which no common structure whatever was present: each simulated participant was assigned an independent random permutation of the same forced quasi-normal distribution of Q = 40 items over nine columns ranging from − 4 to + 4. Person correlation matrices were formed as in Eq. (1) and their eigenvalues

computed, with 2,000 replications at each value of P. Table 2 reports the average number of eigenvalues exceeding unity, the average observed rank and the average and 95th percentile of the leading eigenvalue. Monte Carlo standard errors do not exceed 0.017 for the counts and 0.008 for the leading eigenvalue, so the entries are precise to the digits shown.

Table 2
Eigenvalues of person correlation matrices computed from structureless Q sorts (Q = 40; 2,000 replications per condition)

P	Max. rank	Observed rank	Mean no. of eigenvalues > 1	Mean λ_1	95th percentile of λ_1
20	20	20.0	8.63	2.54	2.85
40	39	39.0	15.76	3.60	3.99
80	39	39.0	27.26	5.41	5.88
100	39	39.0	31.75	6.23	6.71
200	39	39.0	39.00	10.02	10.64

Each simulated Q sort is an independent random permutation of the same forced distribution, so that no common factor exists by construction. Code to reproduce the table in R and in Stata/Mata is available from the author.

Two results deserve emphasis. First, the observed rank equals min(P, Q − 1) in every condition, confirming that the bound in Eq. (2) is attained rather than merely respected. Second, and more consequentially, the Kaiser criterion applied to data containing no structure at all retains 8.6 components on average when P = 20, 31.8 when P = 100 and all 39 extractable components when P = 200. The criterion does not merely become permissive as P/Q grows; in the large-P region it ceases to discriminate altogether.

The leading eigenvalue behaves in the same way. Under the null it averages 2.54 when P = 20 but 10.02 when P = 200, with a 95th percentile of 10.64. A first eigenvalue of, say, 8 would be strong evidence of common structure in a study of 20 participants and no evidence at all in a study of 200. Any reference value used for factor retention must therefore be indexed to the actual P × Q design and not taken from a table or a convention.

3.3 Procedures that require a non-singular person correlation matrix

Singularity has a further consequence that is easily overlooked in applied work, because the software concerned will often return either an error or, worse, a silently unreliable result. Any procedure that requires R^{-1} is unavailable by construction once P ≥ Q. This rules out maximum-likelihood and generalised least squares factor extraction, Bartlett's test of sphericity, the Kaiser–Meyer–Olkin measure

of sampling adequacy, image and alpha factoring, and regression-based (Thurstone) estimation of factor scores. None of these is a defect of the study; they are simply inapplicable to this geometry.

The procedures on which Q methodology actually relies are unaffected. Principal component extraction and centroid extraction operate on R directly and require no inverse; varimax and judgemental rotation operate on the extracted loadings; and factor arrays are built by weighted averaging of defining sorts rather than by regression, so no inverse is needed at that stage either. In other words, the standard Q analytical pipeline is precisely the part of the multivariate toolkit that survives rank deficiency. The practical instruction is therefore narrow: when P ≥ Q, do not request maximum-likelihood extraction or sampling-adequacy diagnostics, and do not interpret their output if a general-purpose package produces it. In general, since there are excellent packages expressly designed for Q-factor analysis (e.g., Kade, Ken-Q), it is advisable not to use general-purpose package like SPSS or similar.

3.4 Parallel analysis adapted to Q designs

Parallel analysis (Horn 1965) compares observed eigenvalues with those obtained from data having the same dimensions but no common structure, and is in principle well suited to the situation described here, since it automatically accommodates the inflation documented in Table 2. One qualification is essential, however. Standard implementations generate random normal deviates or permute a raw data matrix of the same dimensions. Neither null is correct for a forced-distribution Q study, in which every Q sort is a permutation of one fixed set of scores. The appropriate null is the one used in Sect. 3.2: independent random permutations of the study's own forced distribution across the study's own P and Q. A parallel analysis built on a Gaussian null will give the wrong reference distribution for Q data, and the discrepancy grows with P.

A criterion that simply retains fewer factors is not thereby a better criterion: it must also recover structure when structure is present. To check this, Q sorts were generated from a known number of latent viewpoints. Each of 60 simulated participants was assigned to one of n viewpoints, given a noisy realisation of it, and then rank-matched back onto the same forced distribution of 40 items—which is what a forced Q sort is. Because the null distribution depends only on P, Q and the forced distribution, it was computed once, with 3,000 replications, and applied to 500 planted datasets in each of nine conditions. Table 3 reports how often each criterion recovered the planted number of viewpoints.

Table 3
Recovery of a known number of latent viewpoints (P = 60; Q = 40; 500 planted datasets per condition)

Latent viewpoints	Noise SD	Modal no. retained (parallel analysis)	Proportion recovering the correct number	Mean no. retained (Kaiser)
2	0.75	2	1.000	9.95
3	0.75	3	1.000	10.45
4	0.75	4	0.994	10.89
2	1.00	2	1.000	14.07
3	1.00	3	1.000	14.50
4	1.00	4	0.980	14.90
2	1.50	2	1.000	18.06
3	1.50	3	0.962	18.46
4	1.50	4	0.758	18.80

Parallel analysis uses the 95th percentile of the null distribution of each ordered eigenvalue, with retention stopping at the first component that fails. Noise SD is the standard deviation of the disturbance added to a participant's latent vector before it is rank-matched onto the forced distribution. Computed in R. The core computation — standardising each Q sort, forming the person correlation matrix and extracting its eigenvalues — was verified to agree with an independent Stata/Mata implementation to at least seven significant digits on a fixed test case. The two implementations necessarily draw on different random-number streams, so the Monte Carlo summaries in Tables 2 and 3 are not expected to match digit for digit; only the R results are reported here.

The Q-specific parallel analysis recovers the planted number of viewpoints modally in every condition, and exactly in between 96 and 100% of datasets, the single exception being four viewpoints combined with the highest noise level, where it becomes conservative and recovers the correct number in 76% of cases. The Kaiser criterion retains between 10 and 19 components throughout. More revealing is what its count responds to. It rises steeply with the noise level, from 9.95 to 18.06 as the noise standard deviation moves from 0.75 to 1.50, and barely at all with the quantity it is supposed to estimate: holding noise at 0.75, it retains 9.95 components when two viewpoints are present and 10.89 when four are. In these designs the rule tracks the amount of error in the data rather than the amount of structure in it.

Parallel analysis so constructed, taken together with established Q criteria—the number of defining sorts, Humphrey's rule (that the cross-product of a factor's two highest loadings exceed twice the standard error), theoretical coherence and interpretability—is considerably more defensible than mechanical reliance on the Kaiser rule (Watts and Stenner 2012; Zabala et al. 2018).

4 What additional persons do and do not contribute

4.1 The precision of an individual loading

A second consequence follows from the way person correlations are constructed. In conventional statistical analysis, increasing the number of respondents commonly reduces sampling error. It is tempting to suppose that a Q study with 100 participants yields more precise individual factor loadings than one with 30. For an individual Q sort this is not the relevant relationship.

The conventional approximation for the standard error of a Q factor loading is

$$SE \approx 1 / \sqrt{Q}, \quad (4)$$

so that thresholds for a significant loading are usually taken as $|f| > 1.96/\sqrt{Q}$ at approximately the .05 level and $|f| > 2.58/\sqrt{Q}$ at approximately the .01 level (Brown 1980; Watts and Stenner 2012). With Q = 40 the standard error is approximately .158 and the .01 threshold approximately .408. That threshold is identical whether the study contains 20, 40, 80 or 200 persons. Formally, the precision of a loading is a function of Q and not of P:

$$\partial SE/\partial P = 0, \ \partial SE/\partial Q < 0. \quad (5)$$

The approximation in Eq. (4) treats the Q items as independent observations, which they are not exactly: the ipsative structure of a forced sort constrains each person's scores to a fixed mean and variance, and the effective number of independent observations is Q − 1 rather than Q. The convention is nevertheless standard, and the correction is negligible for the values of Q used in practice. The substantive point is unaffected. Adding another person does not add another observation to an existing person's Q sort, which is one of the clearest reasons why conventional notions of statistical sample size cannot be transferred directly from R to Q methodology.

4.2 Composite reliability of a factor array

Additional persons can nevertheless provide statistical benefit through a different mechanism. Once a factor has been identified, its factor array is constructed from the Q sorts that define it. Additional clean defining sorts provide repeated manifestations of the corresponding viewpoint and increase the reliability of the composite representation.

Using the traditional Q-methodological convention of approximately .80 reliability for an individual Q sort, the composite reliability of a factor defined by n sorts follows the Spearman–Brown form

$$r_f = 0.80n / [1 + (n - 1)0.80]. \quad (6)$$

Two defining sorts give a composite reliability of about .889; three, .923; four, .941; five, .952; seven, .966; and ten, .976. Two qualifications should accompany this calculation. The figure of .80 is a convention inherited from the early literature rather than an estimate obtained in any particular study, and the formula assumes that the n defining sorts are parallel measurements of the same factor, whereas in

practice they differ appreciably in loading magnitude. Bootstrap resampling of the Q sorts provides a less idealised assessment of the stability of both loadings and factor arrays (Zabala and Pascual 2016).

More important for the argument of this note is the shape of Eq. (6). The gain from the first few defining sorts is large and the gain thereafter is small: moving from two defining sorts to five raises composite reliability by about .063, whereas moving from ten to twenty raises it by about .012. Increasing the total P set does not automatically increase factor reliability; it does so only insofar as additional participants supply usable defining sorts, and then only with steeply diminishing returns. A study with 100 persons may contain a factor cleanly defined by three of them, while a study with 30 persons may contain a factor defined by ten.

4.3 Multiplicity in large P sets

Increasing P also increases the number of statistical relationships available for inspection. The number of pairwise person correlations is P(P − 1)/2: with 30 persons there are 435, with 100 persons 4,950 and with 200 persons 19,900. Every one of them continues to be estimated from the same Q statement positions. Likewise, if k factors are extracted the analyst examines Pk loadings; with 100 persons and four factors, 400 loadings.

Factor loadings are not independent significance tests—they are constrained within each sort, and rotation is chosen partly in the light of the data—so a formal multiple-comparison correction should not be applied to them. As an order-of-magnitude heuristic, however, it is instructive to note that if loadings behaved as independent tests at a nominal .01 level, a design with P = 100 and k = 4 would be expected to produce about four loadings above threshold even in the complete absence of structure. The simulation in Sect. 3.2 indicates that the real figure under a forced-distribution null is of a similar order. The underlying point is therefore not merely rhetorical: as P grows, the opportunity to encounter apparently noteworthy correlations and loadings grows, while the information contained in each individual Q sort does not.

This matters most for small factors. In a large P set, a factor defined by two apparently significant sorts deserves scrutiny going well beyond the fact that two loadings have crossed a nominal threshold: the pattern of loadings, the distinguishing statements, theoretical coherence and stability under resampling all bear on whether the factor should be retained. Bootstrap or subsampling analyses are especially informative in such cases (Zabala and Pascual 2016).

5 The geometry of the forced Q sort

The structure becomes clearer still when participants use a common forced or quasi-normal distribution. Every Q sort then has the same mean and the same variance. After centring, all Q-sort vectors lie in the same (Q − 1)-dimensional subspace and, under an identical distribution, have the same Euclidean length.

Geometrically, Q sorts can therefore be conceived as points on a hypersphere embedded in a (Q − 1)-dimensional space, the correlation between two sorts corresponding to the cosine of the angle between

them. Q factor analysis identifies systematic common directions in this space.

Increasing P while Q remains constant does not enlarge the space; it adds more points to the same space. This yields a useful shorthand. Q determines the dimensional, or informational, space within which subjectivity can be expressed through the Q sample. P determines how extensively and how diversely that space is populated by observed Q sorts.

6 Redundancy and the marginal value of an additional participant

The geometric argument also clarifies why an indefinitely increasing P set is not necessarily advantageous. Once the principal configurations relevant to the study have been encountered, additional participants may contribute Q sorts very similar to those already present. In statistical terms they add points in already densely populated regions of the Q-sort space rather than revealing new directions within it.

Such redundancy is not intrinsically harmful. Replication strengthens confidence that a factor is not an idiosyncratic one-person configuration, supplies additional clean defining sorts and permits examination of within-factor variation. But the marginal informational gain from an additional participant can decline sharply, and it declines along the same curve as the composite reliability in Eq. (6). A twentieth near-duplicate defining sort adds little to the substantive understanding of a factor, whereas the first participant expressing a previously unobserved viewpoint may add a great deal.

Redundancy also has an analytical consequence. If one type of viewpoint is deliberately or accidentally represented by many highly similar persons, it contributes more heavily to the correlation matrix and tends to produce a larger eigenvalue. A viewpoint represented by 40 closely similar sorts may dominate the extracted variance relative to an equally coherent viewpoint represented by five.

This must not be read as evidence that the first viewpoint is intrinsically more important or more prevalent in the wider population. In a purposively constructed P set, the relative frequency with which a viewpoint has been sampled is in part a design decision. Eigenvalue magnitudes and numbers of defining sorts are therefore conditional on P-set composition (Zabala et al. 2018).

7 Saturation, diversity and purposive coverage

7.1 Saturation as informational redundancy

The language of saturation is attractive here, but it requires qualification. In qualitative research, saturation commonly denotes a point at which further data collection yields little or no new thematic information (Saunders et al. 2018). Something analogous can occur in Q: successive participants may cease to produce meaningfully novel configurations and instead reproduce viewpoints already encountered.

Two senses of the term should nevertheless be kept apart. Saturation of sorts obtains when additional participants no longer occupy new regions of the Q-sort space; saturation of factors obtains when additional participants no longer change the retained factor solution. The two do not coincide. A set of near-duplicate sorts added to an existing cluster can leave the factor solution untouched while visibly increasing the density of one region, and, conversely, a handful of sorts in a sparsely occupied region can alter the solution materially. Only the second sense bears directly on analytical decisions, and it can be examined empirically by re-estimating the solution on subsamples or bootstrap replicates (Zabala and Pascual 2016).

In neither sense should saturation be treated as a strict statistical stopping rule for Q sampling. Q methodology does not attempt to enumerate all possible subjective positions in a population, and an apparently saturated P set cannot demonstrate that no additional factor would emerge if a differently positioned person were recruited. Whether a viewpoint appears depends on who is invited into the P set, on the Q sample, on the condition of instruction and on the social and theoretical boundaries of the study.

A more appropriate aim is therefore diversity, or coverage of theoretically relevant subject positions, rather than saturation in the strong sense. During iterative recruitment, redundancy can still serve as a useful diagnostic: if new participants repeatedly produce configurations already well represented and no longer extend the interpretative range of the study, this may indicate diminishing returns from further recruitment within that particular stratum. The logic is closer to information power—adequacy depending on the relevance and specificity of the sample to the research aim—than to a numerical stopping threshold (Malterud et al. 2016). Saturation may thus be useful as a pragmatic concept of informational redundancy, but not as proof of exhaustive coverage. A better question than ‘has the P set reached saturation?’ is ‘have we purposively covered the theoretically relevant diversity of positions, and are further participants likely to add substantively different configurations?’

7.2 Variety, purposive sampling and representativeness

This distinction returns P-set construction to its methodological purpose. Q methodology ordinarily seeks to reveal the different ways in which a topic can be understood, evaluated or experienced. The objective of P-set construction is therefore not demographic or statistical representativeness of a target population but the purposive inclusion of persons likely to express relevant variation in viewpoint (Previte et al. 2007; Zabala et al. 2018).

A large random or convenience sample can contain substantial numerical redundancy while still omitting an important minority or specialist perspective. Conversely, a much smaller purposive P set can be methodologically richer if recruitment deliberately spans theoretically relevant contrasts—different roles, experiences, institutional positions, levels of involvement or known discursive positions (Palinkas et al. 2015; Previte et al. 2007).

Variety should not be confused with demographic heterogeneity for its own sake. Characteristics such as age, sex, occupation, geography or stakeholder category matter to P-set construction only where there is substantive reason to expect them to expose relevant differences in perspective. The sampling logic is theoretical and purposive: participants are selected because of the positions from which they may engage with the concourse, not in order to reproduce population margins.

This is also why increasing P cannot convert a Q study into a representative survey. If 60 of 100 purposively recruited participants load on Factor A and 10 on Factor B, it does not follow that 60% of the population holds A and 10% holds B. The denominator is a constructed P set, not a probability sample. The relative number of persons loading on each factor is descriptive of the study sample, and may illuminate how the factor solution was formed, but it is not an estimator of population prevalence (Zabala et al. 2018).

Indeed, purposive over-sampling of positions believed to be unusual can be desirable in Q precisely because it increases the chance that a distinctive viewpoint will become visible. Such a strategy is almost the opposite of sampling for representativeness. A rare but theoretically consequential viewpoint may be as important to identify as a common one (Capdevila and Lazard 2009; Zabala et al. 2018).

8 Four quantities that should not be conflated

The preceding considerations suggest separating four quantities that discussions of sample size in Q routinely run together:

- Q, the information available for estimating similarity among Q sorts, and hence the dimensionality of the space within which subjectivity can be expressed;
- P, the opportunity to encounter different configurations of subjectivity;
- P□, the number of defining sorts contributing to the estimation of the array for factor x; and
- the marginal diversity contributed by an additional member of the P set, which two P sets of equal size may possess in very different degrees.

These quantities answer different questions. Q concerns the expressive and correlational basis of the analysis. P concerns exposure to potential viewpoints. P□ concerns the empirical support for a particular factor array. Marginal diversity concerns whether the recruitment strategy has populated the Q-sort space in ways capable of revealing the subjectivities relevant to the research question. Only the first two are fixed by the design; the last two are properties of the data that emerge from it.

9 Implications for practice

Nothing in the foregoing provides a mathematical justification for requiring P < Q. If 100 strategically selected participants sort 40 well-constructed statements, there is no reason to discard participants merely to satisfy an observation-to-variable rule borrowed from R factor analysis.

At the same time, ‘more participants’ should not be treated as an unconditional quality criterion. Five practical implications follow.

- Recognise the rank deficiency. When P ≥ Q, the person correlation matrix is singular by construction and at most Q − 1 factors can be extracted. This is a property of the design, not a fault to be corrected.
- Abandon the Kaiser criterion in large-P designs. Table 2 shows that with Q = 40 and P = 200 the eigenvalue-greater-than-one rule retains every extractable component from data containing no structure at all. Report a scree inspection, a parallel analysis constructed on permutations of the study's own forced distribution, Humphrey's rule, the number and cleanliness of defining sorts, substantive interpretability and, where feasible, bootstrap assessment of stability (Horn 1965; Zabala and Pascual 2016; Zabala et al. 2018).
- Avoid procedures that require R^{-1}. Maximum-likelihood extraction, Bartlett's test, the Kaiser–Meyer–Olkin index and regression-based factor scores are not merely inadvisable but undefined when P ≥ Q. Principal component and centroid extraction remain appropriate.
- Do not expect large P to sharpen individual loadings. The precision of a loading is governed by Q. Where greater precision is needed, lengthen or improve the Q sample rather than recruiting further participants.
- Design the P set for purposive heterogeneity rather than numerical representativeness. Recruitment should maximise theoretically relevant variety, including positions that may be uncommon but important. Additional recruitment becomes less valuable when it merely reproduces already dense regions of the Q-sort space, unless replication is itself required for a particular methodological purpose (Palinkas et al. 2015; Previte et al. 2007; Zabala et al. 2018).

Saturation can inform the last of these judgements if understood modestly, as evidence of diminishing informational returns within a purposive recruitment strategy. It should not be presented as proof that all possible subjectivities have been discovered (Malterud et al. 2016; Saunders et al. 2018).

10 Conclusion

The question of whether a Q study may contain more persons than statements has a simple answer but a less simple explanation. There is no requirement that the P set be smaller than the Q sample. A design in which P > Q is mathematically legitimate and may be methodologically desirable where a broad P set is needed to expose heterogeneous viewpoints.

P and Q are not, however, interchangeable dimensions. Once P exceeds Q the person correlation matrix necessarily becomes singular; its rank is bounded by Q − 1; its non-zero eigenvalues are mechanically elevated, to the point where the Kaiser criterion retains every extractable component from structureless data; an identifiable set of standard multivariate procedures becomes undefined; and the multiplication of correlations and loadings creates further opportunities for apparently noteworthy relationships. None of this invalidates the design, but all of it bears on how the design should be analysed and reported.

Increasing P does not reduce the uncertainty of an individual's factor loading. Its benefits lie elsewhere: in the opportunity to encounter relevant subjectivities and, when additional participants become defining sorts, in the improved representation of particular factors. Beyond some point, additional participants mainly create redundancy. Repeatedly sampling the same viewpoint can strengthen a factor, but it need not increase the diversity of subjectivity revealed by the study, and it can distort the relative magnitude of eigenvalues.

The construction of a P set is therefore better understood as a problem of purposive coverage than of statistical representativeness. The aim is not to reproduce the population numerically but to create sufficient opportunity for theoretically relevant differences in subjectivity to become manifest. Saturation may serve as a pragmatic indicator of diminishing returns, but it should not be elevated into a formal stopping rule or a claim of exhaustive coverage.

The conventional R-technique question—'do we have enough observations for this number of variables?'—is thus not the appropriate question for Q. The useful questions are whether the Q sample provides an adequate expressive space, whether the P set purposively covers the relevant diversity of positions, whether additional participants are revealing new configurations or mainly reproducing existing ones, and whether the resulting factors are supported by enough coherent defining sorts to permit reliable interpretation.

Seen in this way, P > Q is neither a methodological defect nor a matter of complete statistical indifference. It is a particular geometry of Q data, whose mathematical consequences need to be understood alongside the purposive logic by which persons are selected.

Declarations

Competing interests

The author has no competing interests to declare that are relevant to the content of this article.

Ethics approval

Not applicable. This article does not contain any studies with human participants or animals performed by the author.

Use of AI-assisted technologies

The author used ChatGPT (OpenAI, GPT-5.6) during the preparation of this manuscript for language editing and formatting the manuscript to the journal's requirements. Claude (Anthropic) was used for drafting the simulation code reported in Sect. 3.2, and as an interlocutor in developing and reviewing the arguments presented in Sects. 3.1 and 3.3. All analytical results were derived and verified by the author, who reviewed and edited all output and takes full responsibility for the content of this publication.

Funding

The author declares that no funds, grants or other support were received during the preparation of this manuscript.

Author Contribution

The author confirms sole responsibility for the conception of the study, the analysis and the preparation of the manuscript.

Data Availability

The simulation reported in Sect. 3.2 uses no empirical data. Code to reproduce Table 2, implemented independently in R and Stata/Mata, together with a function implementing the Q-specific parallel analysis described in Sect. 3.4, is available at https://doi.org/10.5281/zenodo.22885494

References

1. Brown, S.R.: Political Subjectivity: Applications of Q Methodology in Political Science. Yale University Press, New Haven (1980)
2. Capdevila, R., Lazard, L.: Is it just me… Q methodology and representing the marginal. Oper. Subj. **32**(1) (2009). https://doi.org/10.22488/okstate.08.100567
3. Horn, J.L.: A rationale and test for the number of factors in factor analysis. Psychometrika. **30**(2), 179–185 (1965). https://doi.org/10.1007/BF02289447
4. Kaiser, H.F.: The application of electronic computers to factor analysis. Educ. Psychol. Meas. **20**(1), 141–151 (1960). https://doi.org/10.1177/001316446002000116
5. Malterud, K., Siersma, V.D., Guassora, A.D.: Sample size in qualitative interview studies: guided by information power. Qual. Health Res. **26**(13), 1753–1760 (2016). https://doi.org/10.1177/1049732315617444
6. McKeown, B., Thomas, D.B.: Q Methodology, 2nd edn. Sage, Thousand Oaks (2013)
7. Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K.: Adm. Policy Ment Health. **42**(5), 533–544 (2015). https://doi.org/10.1007/s10488-013-0528-y Purposeful sampling for qualitative data collection and analysis in mixed method implementation research
8. Previte, J., Pini, B., Haslam-McKenzie, F.: Q methodology and rural research. Sociol. Rural. **47**(2), 135–147 (2007). https://doi.org/10.1111/j.1467-9523.2007.00433.x
9. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., Jinks, C.: Saturation in qualitative research: exploring its conceptualization and operationalization. Qual. Quant. **52**(4), 1893–1907 (2018). https://doi.org/10.1007/s11135-017-0574-8

10. Stephenson, W.: The Study of Behavior: Q-Technique and Its Methodology. University of Chicago Press, Chicago (1953)
11. Watts, S., Stenner, P.: Doing Q Methodological Research: Theory, Method and Interpretation. Sage, London (2012)
12. Zabala, A., Pascual, U.: Bootstrapping Q methodology to improve the understanding of human perspectives. PLoS ONE. **11**(2) (2016). https://doi.org/10.1371/journal.pone.0148087 e0148087
13. Zabala, A., Sandbrook, C., Mukherjee, N.: When and how to use Q methodology to understand perspectives in conservation research. Conserv. Biol. **32**(5), 1185–1194 (2018). https://doi.org/10.1111/cobi.13123