Making Cluster Analysis Inferential: From Discovery to Measurement

Abstract

Cluster analysis is political science's primary tool for unsupervised discovery, yet the discipline has never required it to meet inferential standards. This paper introduces the Inferential Cluster Analysis (ICA) framework, which transforms cluster analysis from a descriptive tool into one capable of testing theoretical predictions, validating latent constructs, and supporting genuine inference. The framework builds three conditions into four steps : falsifiability, model fit, and robustness are enforced through theoretical pre-specification, multi-algorithmic estimation, model fitting and selection, and validation. An application tests a foundational assumption in acculturation research that has organized the field for decades. The binary model, which presumes a zero-sum relationship between heritage and American cultural attachments, fails to hold. In its place, four distinct orientations emerge: culture-affirming, assimilationist, bicultural, and demicultural. The framework thus both falsifies a long-standing assumption and validates a novel bidirectional model of acculturation.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Making Cluster Analysis Inferential: From Discovery to Measurement

Machine learning has become one of the most consequential methodological developments in political science, expanding what the discipline can discover, measure, explain, and predict (Grimmer, Roberts, and Stewart 2022). The methods driving that transformation are growing more sophisticated in both design and application, and they are almost entirely supervised.

The discipline's investment in making those methods inferentially credible has been substantial. Supervised learning now operates within a well developed apparatus of identification strategies, cross validation norms, and robustness standards that transform analytic outputs into defensible claims. Unsupervised learning was left outside that project. Cluster analysis, the most widely used unsupervised method in political science, defines no outcome variable and imposes no assumed structure on the relationships among variables, letting the data reveal latent groupings from the predictors alone[1]. It is used to detect meaningful groupings, generate typologies, and explore data before committing to a specific design. Grimmer, Roberts, and Stewart (2022) organize machine learning's contributions to political science around discovery and measurement, and cluster analysis is the discipline's primary tool for discovery. That

[1]Factor analysis and principal component analysis are superficially similar to cluster analysis in that both recover latent structure without a pre-specified outcome variable, but they address a fundamentally different theoretical question. Both methods decompose observed variables into continuous latent dimensions, recovering axes of co-variation rather than discrete groupings. They are appropriate when the question concerns how variables pattern together along underlying continua. When the question is whether distinct types of cases exist in the population, dimensional solutions cannot answer it. Latent class analysis is the closest parametric alternative, recovering latent categorical structure from observed indicators without a pre-specified outcome variable. The distinction between LCA and cluster analysis, and therefore between LCA and ICA, lies in the kind of commitment each requires before the analysis runs. LCA requires the analyst to specify a distributional response function and to assume conditional independence among observed variables within classes (Goodman 1974; Lazarsfeld and Henry 1968). These are structural commitments built into the model that constrain what latent structure it can recover. ICA requires theoretical expectations before analysis runs, but those expectations do not constrain the algorithm. Its commitments are epistemic rather than structural, meaning the data can genuinely falsify the expectation rather than being partially shaped by it. LCA is the appropriate tool once categorical structure has been established through discovery methods like ICA and the analyst is ready to estimate it parametrically with the inferential benefits a likelihood-based framework provides.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

position in the analytic workflow has kept it at the discovery stage, valued for exploration but not asked to test theoretical claims or support genuine inference. This paper closes that gap.

Closing it requires building inferential standards directly into the design. Those standards include theoretical predictions stated before analysis, evidence that findings hold across different modeling assumptions, and confirmation that results are stable and robust. Standard cluster analysis alone cannot meet these requirements. The Inferential Cluster Analysis (ICA) framework builds them directly into the design, transforming cluster analysis from a purely descriptive tool into one capable of testing theoretical claims, validating latent constructs, and supporting genuine inference.

The framework rests on three conditions that any inferential design for unsupervised learning must satisfy. Falsifiability requires that theoretical expectations be stated in a form the data can contradict before any analysis runs. Model fit requires that identified structures hold across fundamentally different algorithmic approaches rather than reflect the assumptions of any single method. Robustness requires that findings survive systematic scrutiny under resampling and cross-validation. The ICA framework builds these conditions directly through four integrated steps: theoretical pre-specification, multi-algorithmic estimation, model fitting and selection, and validation.

The paper proceeds in four parts. The first section establishes what moving cluster analysis from description to inference requires, specifying three conditions any inferential design must satisfy. The second introduces the four steps of the ICA framework, showing how each maps onto a specific condition. The third demonstrates the framework in practice through a replication of Author (under review), which uses ICA to test a foundational assumption about acculturation that has organized two decades of political science research without ever being

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

directly examined. That application shows what the framework makes possible. A long-standing assumption is falsified, a novel theoretical model is validated, and a previously unknown orientation is discovered, all within a single pre-specified design. The fourth discusses the framework's scope conditions, limitations, and opportunities for extension.

From Description to Inference

King, Keohane, and Verba (1994) define inference as using observations from the world to learn about what has not been observed, and that definition draws no boundary around method. Making cluster analysis inferential requires that three conditions be met, each following directly from the unsupervised nature of the method, and each addressing a distinct vulnerability that standard cluster analysis cannot meet.

The first condition is falsifiability. A claim is inferential only if it can in principle be contradicted by evidence, and that requires stating theoretical predictions in a form the data can disconfirm before analysis begins (Popper 1959). For cluster analysis, this condition carries particular force. The method specifies no outcome variable and imposes minimal assumptions about functional form, relationships among variables, or error structure. That flexibility leaves no internal benchmark against which results can be evaluated. Theoretical predictions supplied before analysis fill that gap. They provide a standard the data can meet or fail to meet, transforming discovery into test. Pre-registration is the most direct way to enforce this commitment, though any documented pre-analysis plan that demonstrably precedes the analysis satisfies the condition (Gerber and Green 2012; Humphreys, Sanchez de la Sierra, and van der Windt 2013). What matters is that theoretical predictions exist before results do, so that confirmation, disconfirmation, and discovery are all possible outcomes and each carries inferential weight.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

The second condition is model fit. Because cluster analysis specifies no outcome variable, there is no a priori basis for knowing which algorithm best captures the underlying structure in the population. Every clustering algorithm imposes a distinct set of modeling assumptions, and those assumptions determine what structure the algorithm can and cannot find. Running only one algorithm leaves no way to know whether the result reflects genuine structure or merely those assumptions. The model fit condition addresses this by requiring a comparative approach. Multiple algorithms with fundamentally different mathematical foundations are applied to the same data, and the pattern of results across them becomes the basis for understanding what structure the data contain. K-means partitions cases by minimizing within-cluster variance around centroids, producing solutions shaped by Euclidean distance geometry that tend toward spherical, roughly equal-sized clusters (MacQueen 1967). Gaussian Mixture Models assign probabilistic membership based on distributional assumptions about how the data were generated, accommodating elliptical cluster shapes and varying covariance structures that centroid-based methods cannot represent (McLachlan and Peel 2000). Agglomerative clustering builds hierarchical structure from pairwise similarities among cases, operating independently of both distance geometry and distributional assumptions, so that the solution reflects local similarity structure rather than any assumed global geometry (Ward 1963). No single algorithm is privileged. The question is not whether structure holds across all of them, though convergence when it occurs provides strong evidence. The question is what the comparative performance reveals. Some algorithms will fit better than others. Some will recover structure that others miss. Divergence is itself diagnostic, telling the analyst that the data contain features that some algorithms can represent and others cannot. The model fit condition is satisfied when the analyst

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

uses this comparative information, together with statistical performance metrics, cluster balance, and theoretical interpretability, to select the solution that best captures what the data contain.

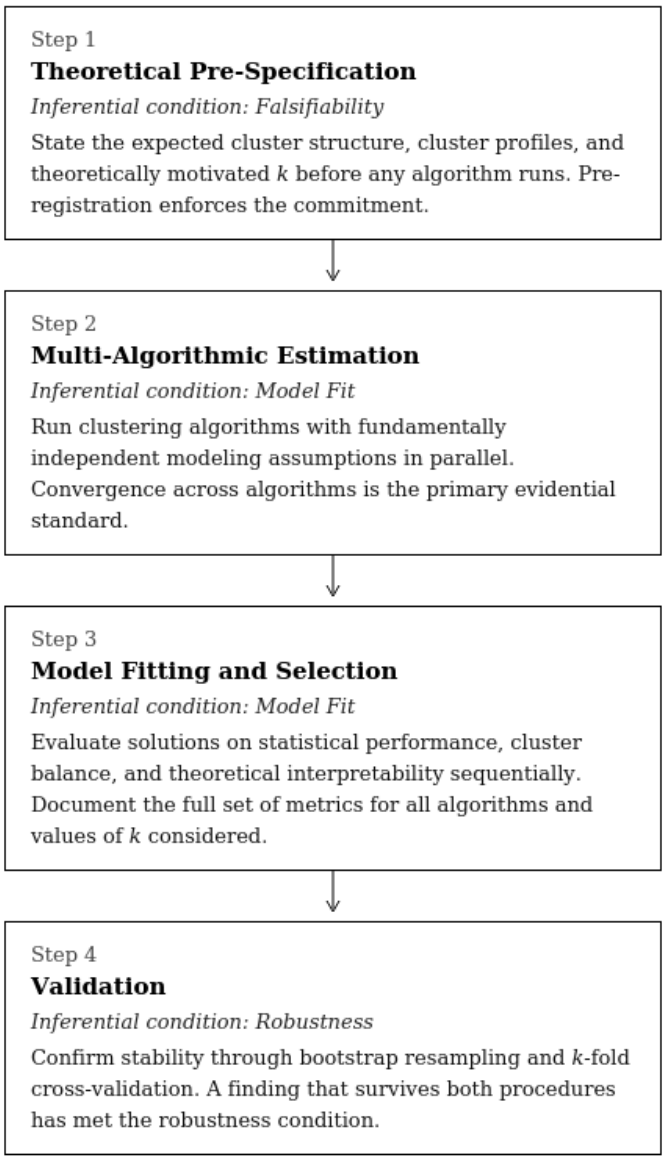
Robustness is the third condition. A finding can satisfy the first two conditions, emerging from pre-specified theoretical predictions and performing well across multiple algorithms, yet still be fragile. It may be an artifact of this particular sample, a pattern that emerged because of which observations happened to be included and would not reproduce if the data were collected again. It may also be contingent on specific analytical choices within the chosen algorithm, threshold decisions or convergence criteria that could have been set differently. Robustness requires that the finding survive systematic scrutiny designed to expose such fragility. The claim is not yet about generalizing to populations beyond the one studied, though that may come later. The claim is more fundamental. The finding would reproduce under resampling and across different subsets of the data, meaning it is not an accident of this sample or these analytical decisions (Efron 1979). Political science has increasingly demanded this standard of empirical findings across methods, and there is no principled basis for exempting cluster analysis from it. Bootstrapping and k-fold cross-validation address this condition directly. Bootstrapping tests whether the result depends on the specific set of observations, generating resampled datasets and reapplying the clustering procedure to each. Consistent performance metrics and narrow confidence intervals across iterations indicate that the result is not driven by any particular subset of cases. K-fold cross-validation tests something distinct, evaluating whether the solution holds across different portions of the data by partitioning the dataset and assessing stability in each held-out fold. Together these procedures provide systematic evidence that the finding is robust rather than artifact, stable rather than contingent.

The Inferential Cluster Analysis Framework

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

The three inferential conditions established above, falsifiability, model fit, and robustness, define what any inferential design for cluster analysis must achieve. The ICA framework is built around four steps that map directly onto those conditions, each one addressing what the others cannot, and together constituting a complete inferential design for unsupervised learning, as shown in Figure 1.

Figure 1. The Four Steps of the Inferential Cluster Analysis Framework



Step 1. Theoretical Pre-Specification

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Theoretical pre-specification requires the analyst to state, before any algorithm runs, what structure theory leads them to expect in the data. Those expectations must be specific and falsifiable, committing to how many clusters should emerge, where those clusters should be positioned relative to each other in variable space, and what profiles they should exhibit across the variables used to identify them. The goal is not to predict the result with certainty but to establish a standard against which the result can be evaluated, transforming what would otherwise be open-ended exploration into a genuine test.

The most critical parameter in any cluster analysis is the number of clusters *k*, and the ICA framework treats this as a theoretical commitment rather than a technical decision made after seeing the data. In addition to specifying expectations about cluster profiles and patterns, the analyst should theorize how many clusters are expected to emerge and why, grounding that expectation in the same theoretical reasoning that motivates the analysis. If theory predicts four distinct groupings in the population, *k* equals four is the theoretically motivated starting point, and that choice should be stated and defended in advance. Pre-registration provides a documented record that theoretical expectations preceded empirical results. Platforms such as OSF offer one straightforward way to do this, though any pre-analysis plan that demonstrably precedes the analysis satisfies the condition (Gerber and Green 2012; Humphreys, Sanchez de la Sierra, and van der Windt 2013). Pre-registration is recommended rather than required because institutional and data contexts vary, but the underlying principle of pre-specifying one's approach is not negotiable.

Without a prior commitment, any cluster solution can be made to appear theoretically meaningful through post-hoc reasoning, because the analyst retains full interpretive freedom after seeing the results. Pre-specification removes that freedom selectively. Unexpected findings can

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

and do emerge within a pre-specified design, and the framework accommodates them. Pre-specification does not prevent discovery, rather, it makes discovery interpretable as discovery rather than post-hoc invention, because the theoretical interpretation of any result is constrained by what was committed to beforehand. This applies to the choice of k as much as to the expected cluster profiles, since selecting the number of clusters after seeing which value produces the cleanest solution is the same interpretive problem pre-specification is designed to prevent. A critical advantage of this approach is that confirmation, revision, and genuine discovery are all possible outcomes, and each carries inferential weight precisely because the theoretical commitment was made before the results were known. The framework is theory-driven in its expectations and data-driven in its results, and what the data reveal, whether it aligns with those expectations or departs from them, is meaningful because the standard against which it is evaluated was established in advance.

Step 2. Multi-Algorithmic Estimation

Multi-algorithmic estimation requires the analyst to run multiple clustering algorithms that differ fundamentally in their mathematical assumptions and to treat the pattern of results across those algorithms as the primary basis for understanding what structure the data contain. A single algorithm, no matter how well executed, can only ever tell the analyst what its specific assumptions reveal. Running multiple algorithms with genuinely independent modeling assumptions provides something no single algorithm can, the ability to see whether a finding persists across fundamentally different ways of constructing clusters or whether it depends on a particular set of mathematical commitments.

Before selecting algorithms, the analyst should consider the range of approaches available and what each is designed to do. Clustering algorithms differ not only in their

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

mathematical foundations but in what kinds of structure they are built to find. Hard clustering algorithms, such as *k*-means and agglomerative methods, assign each case exclusively to one cluster, which is appropriate when the theoretical expectation is that groups are relatively distinct (Kaufman and Rousseeuw 1990). Soft clustering algorithms, such as Gaussian Mixture Models and Fuzzy C-Means, assign probabilistic membership across clusters, accommodating the possibility that cases sit at the boundaries between groups, which matters when theory predicts overlapping or hybrid orientations rather than clean categorical distinctions (Bezdek 1981). Density-based approaches such as DBSCAN take a different view entirely, identifying clusters as regions of high density separated by regions of low density and treating cases in sparse regions as noise rather than forcing them into a cluster, which is valuable when the analyst suspects that outliers or data irregularities might otherwise distort the solution (Ester et al. 1996). The choice among these approaches is not arbitrary. It should be guided by theoretical expectations about the nature of the structure being sought, and that choice should itself be pre-specified and justified, because the rationale for algorithm selection is part of the analytical commitment the framework requires.

Multi-algorithmic estimation shifts the question. The goal is no longer to find the single best fitting algorithm but to understand what structure emerges across fundamentally different ways of modeling the data. A finding that emerges from only one algorithm could reflect that algorithm's assumptions rather than genuine structure in the population. The results across algorithms are therefore best treated as diagnostic information. Where algorithms agree and where they disagree each tell the analyst something meaningful about the underlying structure in the data. Full convergence, where independent algorithms recover the same structure, provides the strongest evidence that the identified clusters reflect something real rather than a methodological artifact.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Partial convergence, where algorithms agree on the broad structure but differ in compactness or boundary definition, is itself informative about the nature of the clusters and should be interpreted alongside statistical metrics rather than dismissed. Full divergence signals that the variables may be poorly suited to the theoretical question, that the assumed structure does not exist, or that the sample lacks sufficient signal to recover it reliably. The principle that convergence across independent methods strengthens evidential claims has deep precedent in the validation literature (Campbell and Fiske 1959; Adcock and Collier 2001). Multi-algorithmic estimation applies that principle directly to the problem of identifying latent structure, where no single algorithm's output is sufficient on its own.

Step 3. Model Fitting and Selection

Multi-algorithmic estimation reveals what structure the data contain across different modeling assumptions, but it does not by itself determine which solution best captures that structure or whether the pre-specified k is the right one. Model fitting and selection answers those questions by giving the analyst a clear, ordered basis for evaluating what the algorithms produced and arriving at a defensible choice. Because k was committed to in Step 1, the analyst enters this step with a primary candidate, but examining performance at k-1 and k+1 is an important check. A well-supported solution should consistently outperform neighboring values across algorithms and metrics, and any solution that performs comparably or better warrants a closer look before the selection is finalized. The selection decision rests on three criteria that must be integrated rather than applied mechanically: statistical performance, cluster balance, and theoretical interpretability. Each contributes something the others cannot, and none is sufficient on its own.

First, statistical performance is assessed through two metrics that evaluate different dimensions of cluster quality. The first is the silhouette score, which measures how similar each

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

case is to its assigned cluster relative to the nearest competing cluster, producing a value that ranges from zero to one. A score approaching one indicates that cases are well-matched to their assigned cluster, while lower scores suggest weaker boundaries and greater overlap between clusters. Scores below 0.25 suggest weak structure and warrant careful consideration before accepting the solution (Rousseeuw 1987). The appropriate threshold depends on the nature of the constructs being measured, since constructs with theoretically overlapping boundaries will naturally produce lower scores, and researchers should interpret them in light of their theoretical expectations rather than against a fixed cutoff.

The second metric is the Calinski-Harabasz index, which evaluates the ratio of between-cluster dispersion to within-cluster dispersion, with higher values indicating denser and better-separated clusters (Calinski and Harabasz 1974). Unlike the silhouette score, which evaluates how well individual cases fit their assigned cluster, the Calinski-Harabasz index evaluates how distinct the clusters are from each other relative to how compact they are internally, making it particularly useful for comparing solutions across different values of k on a common scale. Together these two metrics provide complementary assessments of solution quality, and neither should be interpreted in isolation since each captures a different dimension of cluster structure and each can be misleading when read alone.

The second criterion is cluster balance, meaning the distribution of cases across clusters. Solutions in which one cluster contains nearly all observations while others contain very few raise two distinct concerns. The first is substantive, because a cluster with very few members may not represent a meaningful or stable grouping in the population. The second is statistical, because small clusters produce unreliable estimates of cluster profiles and lack sufficient power for downstream analysis. As a practical guideline, the framework recommends a minimum cluster size of

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

approximately five percent of the total sample, though analysts should adjust this threshold based on total sample size and the demands of subsequent analysis. Solutions that produce clusters falling below this threshold should be examined carefully before being accepted, and the rationale for accepting or rejecting them should be documented transparently.

The third criterion is theoretical interpretability, which is distinct from theoretical confirmation and requires careful handling. A cluster solution is theoretically interpretable when the profiles of the recovered clusters can be meaningfully connected to the theoretical expectations stated in pre-specification, whether those clusters match the expected profiles, depart from them in informative ways, or reveal structure the theory did not anticipate. Theoretical interpretability does not require that every pre-specified cluster emerge exactly as predicted. It requires that the analyst be able to give a theoretically coherent account of what each cluster represents and why, an account grounded in the pre-specified expectations rather than constructed freely after the fact. Discovery is not only possible within this criterion but expected, because the unsupervised nature of cluster analysis means that structure the theory did not fully anticipate can and does emerge. The interpretability criterion asks whether that structure can be understood in relation to the theoretical framework, not whether it was predicted in advance.

Model selection integrates these three criteria sequentially. Statistical metrics identify the candidate solutions and algorithms that perform well across multiple dimensions of cluster quality. Cluster balance eliminates solutions that are statistically implausible or analytically insufficient. Theoretical interpretability determines which among the remaining candidates best advances the theoretical argument. The final selection should be documented with the full set of metrics reported for all algorithms and all values of k considered, so that the basis for the decision is transparent and reproducible.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Step 4. Validation

A solution that survives model fitting and selection has met three of the four steps of the ICA framework, but one question remains open. The solution may perform well statistically, hold across multiple algorithms, and align with theoretical expectations, yet still reflect a configuration of the data that is specific to this sample. Validation addresses that question by testing whether the solution holds under conditions that would expose an unstable result.

Two procedures address this directly and test different dimensions of stability. Bootstrap resampling tests whether the solution holds under variation in the specific observations included in the sample, generating resampled datasets by sampling with replacement from the original data and applying the clustering procedure to each (Efron 1979). Consistent performance metrics and narrow confidence intervals across iterations indicate that the result does not depend on any particular subset of observations. K-fold cross-validation tests something distinct, evaluating whether the solution holds across different portions of the data by partitioning the dataset into k subsets and evaluating performance in each held-out portion (Stone 1974). Five-fold cross-validation is recommended as a practical starting point. Consistent silhouette scores across folds indicate that the solution emerges reliably from different portions of the data rather than from the sample as a whole. A finding that survives both procedures has done more than demonstrate stability. Following Campbell and Fiske (1959), each independent confirmation adds evidential weight to the claim, so that a solution confirmed across multiple algorithms, resampled datasets, and cross-validated folds accumulates the kind of convergent evidence that makes the finding progressively harder to dismiss as a methodological artifact or a property of the sample alone. That accumulation transforms a stable cluster solution into a validated inferential claim about latent population structure.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

The ICA Framework Applied: Testing Acculturation in American Politics

A framework's value is demonstrated not by its design but by what it produces when put to work on a problem with genuine inferential stakes. This section applies the ICA framework to a foundational assumption that has organized two decades of political science research on immigrant incorporation without ever being subjected to direct empirical scrutiny. The question is whether the binary model of acculturation, which treats heritage and American cultural attachments as mutually exclusive, accurately describes how Latinos in the United States experience cultural attachment. Do they experience acculturation as a zero-sum trade-off where stronger attachment to one culture necessarily implies weaker attachment to the other, or can heritage and American attachments coexist independently?

Acculturation research in political science has long rested on this binary assumption, treating heritage and American attachments as opposing poles rather than independent dimensions. By design, this framework excludes the possibility of any blending between cultures, making hybrid orientations theoretically impossible and forcing every individual into one of two predetermined categories, either oriented toward the heritage culture or oriented toward American culture. The framework is largely deterministic, deciding in advance what acculturation can look like before any data are examined. This assumption has never been directly tested, despite determining how acculturation is measured, how it is modeled, and what findings are even possible to produce.

The analysis focuses on Latinos in the United States, a population for whom acculturation is both theoretically central and empirically consequential for understanding political behavior, incorporation, and identity. The study draws on data from the Latino National Survey (Fraga et al. 2006), which measures heritage and American identity as independent dimensions rather than

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

forcing a trade-off between them. That independence is not a minor technical feature. It is what makes the data capable of adjudicating between the binary and bidirectional frameworks, because a survey that forces a trade-off between heritage and American identity cannot, by design, find evidence of biculturalism even if it exists.

Cluster analysis is the right tool for this question. The problem is fundamentally about latent structure, and the expected complexity of that structure, with theoretically overlapping orientations and fuzzy boundaries between groups, requires an unsupervised approach that does not impose categories before the data have been examined. The ICA framework provides the inferential discipline needed to test whether the binary assumption holds or whether a more complex structure emerges.

Pre-Specification

Before any algorithm ran, two competing theoretical predictions were registered. The binary hypothesis predicted that the data would reveal two orientations positioned at opposing cultural poles, heritage-oriented and American-oriented, with no meaningful concentration of cases in hybrid or intermediate positions. The Bidirectional Acculturation Model (BAM), a novel framework introduced in Author (under review), predicted something different. Rather than treating heritage and American identity as opposing poles, the BAM treats them as independent dimensions, each of which can be high or low independently of the other. This independence produces four theoretically distinct orientations defined by the combination of attachments across both dimensions. Bicultural individuals exhibit strong attachment to both cultures simultaneously, directly contradicting what the binary model treats as impossible. Assimilationist individuals exhibit strong American identity alongside weaker heritage attachment. Culture-affirming individuals exhibit strong heritage attachment alongside weaker American identity. Indifferent

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

individuals exhibit weaker attachment to both. The BAM therefore predicts four clusters, making *k*=4 the theoretically motivated starting point. These four orientations, shown in Figure 2, constitute the theoretical fingerprint the analysis was designed to test.

Figure 2. The Bidirectional Acculturation Model

	High Ancestral Identity	**Low Ancestral Identity**
High American Identity	*Bicultural*	*Assimilationist*
Low American Identity	*Culture-Affirming*	*Indifferent*

Note: Cells represent theoretically predicted acculturation orientations. The BAM treats heritage and American identity as independent dimensions, each of which can be high or low independently of the other.

Results

The analysis began by evaluating five clustering algorithms, with three selected for the final multi-algorithmic design. GMM was included because the BAM predicts orientations with inherently overlapping boundaries that hard clustering approaches cannot accommodate, and because its probabilistic membership structure is uniquely consistent with bidirectional theory (McLachlan and Peel 2000). *K*-means and hierarchical clustering were included as hard clustering approaches that bring independent mathematical assumptions to the problem, ensuring that any convergence across all three reflects genuine structure rather than a particular set of mathematical commitments (MacQueen 1967; Ward 1963).

The exclusion of Fuzzy C-Means and DBSCAN is itself informative about the nature of the latent structure. Fuzzy C-Means produced near-equal cluster sizes inconsistent with every other solution and a low fuzzy partition coefficient. This indicates that the orientations in this population

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

have meaningful boundaries rather than the smooth, overlapping gradations that fuzzy methods are designed to find. DBSCAN did not converge to k=4 under any parameter configuration, ruling out the possibility that the structure is density-based and strengthening the case that the four-cluster solution reflects genuine categorical groupings rather than artifacts of algorithmic assumptions. Full diagnostics for all five algorithms are reported in Appendix Table A1.

Three patterns emerged from the analysis across the three selected algorithms. First, no algorithm at any value of k recovered a two-cluster solution consistent with the binary hypothesis, providing clear and immediate evidence against the binary assumption. Second, all three algorithms converged on a four-cluster solution as the best-fitting solution across statistical metrics, cluster balance, and theoretical interpretability. Third, the five-cluster solution consistently underperformed, with one cluster falling below the minimum size threshold and no meaningful theoretical account available for the additional orientation.

With the four-cluster solution established as the best fit, the next question was which algorithm's implementation of that solution to select. The GMM k=4 solution was chosen as optimal because its probabilistic membership structure is uniquely consistent with the BAM's prediction of overlapping orientations, and because it produced the most theoretically interpretable cluster profiles while demonstrating superior stability in validation procedures.

Table 1. Cluster Analysis Performance Comparison

Algorithm	Silhouette	CH Index	Cluster Sizes	Selection Rationale
K-Means	0.39	1458.53	671, 416, 2808, 890	Strong metrics but hard assignments prevent

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

				probabilistic interpretation
Agglomerative	0.38	1353.55	999, 2748, 586, 452	Similar performance but less theoretically interpretable cluster profiles
GMM	**0.36**	**1063.27**	**433, 705, 378, 3269**	**Selected for theoretical fit and probabilistic membership structure**

Note: GMM was selected despite slightly lower cohesion metrics because (1) it allows probabilistic cluster membership consistent with bidirectional theory, (2) it produced the most theoretically interpretable profiles, and (3) it demonstrated superior stability in validation procedures.

Validation

Bootstrap resampling across 1,000 iterations and five *k*-fold cross-validation tested whether the four-cluster solution would hold under conditions designed to expose instability, satisfying the robustness condition introduced earlier. Bootstrapping produced a mean silhouette score of 0.34, nearly identical to the original solution score of 0.36, with a confidence interval of 0.34 to 0.39 indicating that the result does not depend on any particular subset of observations. Five *k*-fold cross-validation produced consistent silhouette scores across folds, confirming that the four-cluster structure emerges reliably from different portions of the data rather than from the sample as a whole. Following Campbell and Fiske (1959), each independent confirmation adds evidential weight to the claim. A solution that holds across three independent algorithms, 1,000 bootstrap iterations, and five cross-validation folds accumulates convergent evidence that makes it progressively harder to dismiss as a methodological artifact or a property of the sample alone.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

That accumulation transforms a stable cluster solution into a validated claim about latent population structure. Full robustness metrics are reported in Appendix Table A2.

Table 2. Metrics for the Optimal *k*=4 GMM Clustering Solution

Metric	Value	Bootstrap (1,000 iter.)	5-Fold Cross-Validation
Silhouette Score	0.36	Mean = 0.34, 95% CI [0.34, 0.39]	Mean = 0.34 across folds
Calinski-Harabasz Index	1063.27	Mean = 1050.12	Mean = 1050.12 across folds
Cluster Sizes	433, 705, 378, 3269	Bootstrap variation: ±50, ±70, ±40, ±200	Stable across all folds

Note: Metrics evaluate k=4 GMM clustering of standardized acculturation variables. While scores indicate only moderate cohesion, convergent performance across three algorithms and bootstrapped confidence intervals confirms that the result is stable rather than an artifact of this particular sample. Source: 2006 Latino National Survey (LNS, n=4,785).

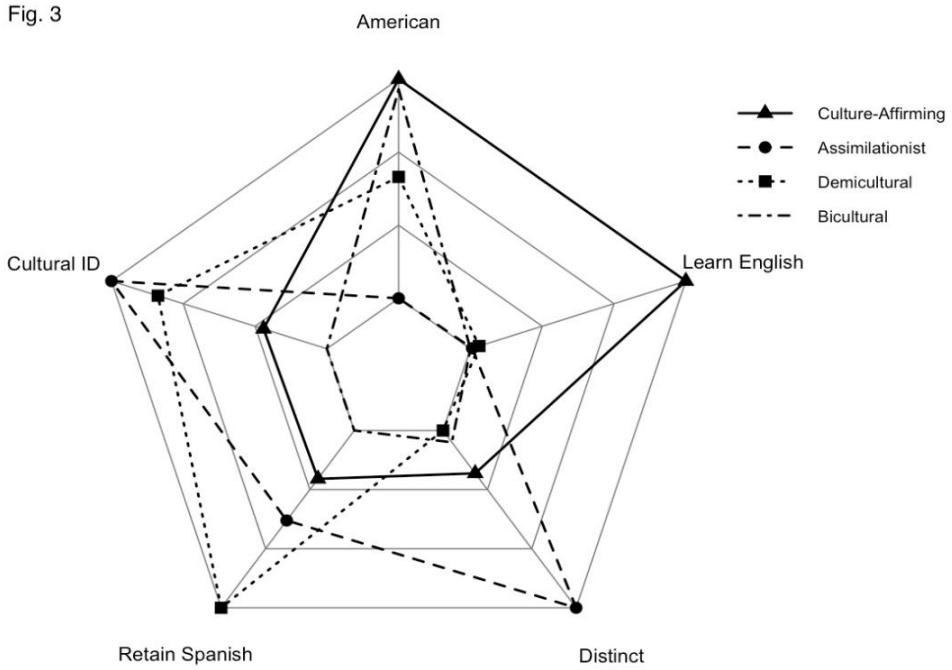
Findings

The four orientations that emerged from the analysis are displayed in Figure 3. Three correspond closely to the theoretical expectations of the BAM. The bicultural orientation, characterized by strong attachment to both heritage and American culture simultaneously, accounts for 68.4% of the sample, making it by far the most prevalent acculturation orientation among Latino adults in the United States. The assimilationist orientation, characterized by stronger American identity alongside weaker heritage attachment, accounts for 14.7%. The culture-affirming orientation, characterized by strong heritage attachment alongside weaker American identity, accounts for 9.0%.

Figure 3. Radar Chart of Cluster Means (*k*=4, GMM)

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Fig. 3



Two findings depart from the pre-specified theoretical expectations in ways that are themselves inferentially significant. The indifferent orientation did not materialize. This is not a null result to be set aside, but it must be interpreted carefully. The absence of an indifferent cluster in this sample could mean that full detachment from both cultures is genuinely rare among Latino adults, that it exists but the measures did not capture it, or that the sample underrepresents the relevant population. What the data show is that the predicted indifferent orientation did not emerge as a distinct cluster under the conditions of this analysis. This refines the BAM and suggests that the theoretical expectation of indifference may require revision. In its place, a fourth orientation emerged consistently across all three algorithms, one the author terms demicultural. Demicultural individuals exhibit moderate attachment to both cultures simultaneously, scoring neither high nor low on either dimension, occupying a middle position that is distinct from the strong dual attachment of bicultural individuals and distinct from the single-culture orientations of assimilationist and culture-affirming individuals. This is not a residual category absorbing cases

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

that did not fit elsewhere. Its consistent emergence across three mathematically independent algorithms, each with fundamentally different modeling assumptions, rules out the explanation that it is a methodological artifact, and because theoretical expectations were committed to before the analysis ran, it is identifiable as a genuine discovery rather than a post-hoc label. The demicultural orientation raises questions that existing acculturation frameworks are not equipped to answer, including what political consequences follow from moderate engagement with both cultures, how demicultural individuals respond to immigration policy appeals, and whether this orientation reflects a transitional stage or a stable position. Each of those is a direction for future research that the ICA framework makes visible precisely because it did not impose categories before examining the data.

The implications of these findings are consequential. Hybrid orientations, bicultural and demicultural combined, account for 76.3% of Latino adults in the sample. Binary acculturation models, which treat hybrid orientations as theoretically impossible, misclassify more than three quarters of Latinos by forcing them into categories that do not reflect how they actually navigate cultural attachment. Two decades of research on Latino political behavior have been organized around an assumption the data directly falsify, with consequences for how acculturation is measured, how its political effects are modeled, and what conclusions the field has drawn about Latino incorporation. The ICA framework made that visible by providing the inferential discipline needed to test what had only been assumed. With the application complete, the final sections turn to the framework's scope conditions, its limitations, and the opportunities it opens for future work.

Scope Conditions, Limitations, and Extensions

The ICA framework applies when three conditions are met. The first condition is theoretical. The research question must concern latent structure in a population, meaning the

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

analyst has reason to believe that meaningfully distinct groupings exist and that theory can generate falsifiable predictions about what those groupings look like before the data are examined. The framework is not designed for purely exploratory work where no theoretical priors exist, and it is not a substitute for causal identification. The second condition is data. The variables used in the analysis must measure the relevant dimensions independently rather than forcing trade-offs between them, because data that pre-imposes structure cannot adjudicate between competing theoretical predictions about what that structure is. The third condition is sample size. The dataset must be large enough to support cluster balance above the recommended threshold across all solutions considered, because solutions that cannot meet the balance requirement lack the statistical foundation the validation procedures require.

When these conditions are satisfied, the framework is applicable across a wide range of substantive domains. Whether the question concerns political identity, public opinion, legislative behavior, or comparative politics, wherever theory predicts meaningful latent groupings that survey or behavioral data can recover, ICA provides a design for testing those predictions.

Even when the scope conditions are met, the framework has limitations, each of which points toward a direct opportunity for refinement. The first is that the framework validates stability but does not yet quantify how precisely the cluster profiles are estimated. While bootstrapping and cross-validation confirm that a solution is stable, neither produces confidence intervals around cluster profile means or membership probabilities, leaving the precision of the inferential claim underspecified in ways that matter for downstream analysis. Future work should develop procedures for quantifying uncertainty around cluster profiles that integrate directly with the validation step.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

The second limitation concerns parameter sensitivity within algorithms. Multi-algorithmic estimation addresses reliance on any single algorithm, but choices such as distance metrics, linkage criteria, and covariance structure assumptions remain at the analyst's discretion. A systematic sensitivity analysis protocol, one that reports whether core findings hold under alternative parameterizations, would provide more principled guidance than current practice offers. Future work should develop such a protocol and integrate it into the framework.

With the framework's conditions, limitations, and extensions now stated, the final section turns to what ICA demonstrates about the possibilities for unsupervised learning in political science.

Cluster Analysis as an Inferential Tool

Grimmer, Roberts, and Stewart (2022) organize machine learning's contributions to political science around discovery and measurement. In their framing, unsupervised methods generate the conceptual raw material that supervised methods then operationalize and test.. The ICA framework demonstrates that cluster analysis, properly designed, can contribute at both stages simultaneously. It can discover latent structure while also testing theoretical claims about what that structure should look like, all within a single pre-specified design.

The acculturation application shows what this makes possible. The binary model had organized two decades of research without ever being tested. ICA made testing possible by building falsifiability, model fit, and robustness directly into the design. The result was both a disconfirmation and a discovery. The binary model failed to hold, and the consequences are stark. It would misclassify more than 75 percent of Latino adults, forcing them into categories that do not reflect how they actually navigate heritage and American cultural attachments. In place of the binary model, a novel four-orientation structure emerged, one that included a previously unknown

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

demicultural category. Hybrid orientations, bicultural and demicultural combined, account for over 76 percent of the sample, revealing that blending, not zero-sum trade-offs, is the dominant experience of acculturation for Latino adults. The framework thus does more than describe what the data contain. It validates a new theoretical typology, refines an existing one, and offers a much-needed correction to a field organized for decades around a false assumption, all within an inferentially credible design.

The same logic applies across political science wherever theory predicts latent groupings that data can recover. Consider three examples. Studies of legislative behavior could use ICA to test whether voting blocs align with predicted partisan or factional divisions, moving beyond descriptive accounts of coalition patterns to actual tests of competing coalition theories. Public opinion research could use it to validate whether issue publics cluster as theories of political psychology expect, transforming what have been post-hoc classifications into confirmatory tests. Comparative politics could use it to examine whether regime types emerge from institutional data in ways that match typological theories, providing a way to adjudicate between competing classification schemes. In each case, ICA offers a design for testing what has been theorized and discovering what has not.

What this paper delivers is not a method but a standard. Cluster analysis has long been treated as a descriptive tool, not because unsupervised learning cannot support inference, but because the discipline has never asked it to meet inferential standards. The ICA framework builds those standards in, and the acculturation application shows what happens when they are met. Assumptions get tested, theories get refined, and latent structure becomes visible. What emerges is a method no longer confined to exploration, one that takes its place alongside supervised learning as a tool capable of supporting genuine inference.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

References

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." American Political Science Review 95(3): 529–546.

Author. Under Review. "Myth of the Zero-Sum: Rethinking Acculturation in American Politics." Politics, Groups, and Identities.

Bezdek, James C. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press.

Calinski, Tadeusz, and Jerzy Harabasz. 1974. "A Dendrite Method for Cluster Analysis." Communications in Statistics 3(1): 1–27.

Campbell, Donald T., and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." Psychological Bulletin 56(2): 81–105.

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." Annals of Statistics 7(1): 1–26.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 226–231.

Fraga, Luis Ricardo, John A. Garcia, Rodney E. Hero, Michael Jones-Correa, Valerie Martinez-Ebers, and Gary M. Segura. 2006. Latino National Survey (LNS), 2006. ICPSR20862-v1.

Gerber, Alan S., and Donald P. Green. 2012. Field Experiments: Design, Analysis, and Interpretation. New York: W. W. Norton.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Goodman, Leo A. 1974. "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models." Biometrika 61(2): 215–231.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton, NJ: Princeton University Press.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." Political Analysis 21(1): 1–20.

Kaufman, Leonard, and Peter J. Rousseeuw. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. New York: John Wiley & Sons.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Lazarsfeld, Paul F., and Neil W. Henry. 1968. Latent Structure Analysis. Boston: Houghton Mifflin.

MacQueen, J. B. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 281–297. Berkeley: University of California Press.

McLachlan, Geoffrey J., and David Peel. 2000. Finite Mixture Models. New York: John Wiley & Sons.

Popper, Karl R. 1959. The Logic of Scientific Discovery. London: Hutchinson.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." Journal of Computational and Applied Mathematics 20: 53–65.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0

Stone, M. 1974. “Cross-Validatory Choice and Assessment of Statistical Predictions.” Journal of the Royal Statistical Society, Series B 36(2): 111–147.

Ward, Joe H., Jr. 1963. “Hierarchical Grouping to Optimize an Objective Function.” Journal of the American Statistical Association 58(301): 236–244.

AI Disclosure: The author used Gemini and Claude for proofreading, grammar checks, and R coding support. All scholarly content, theoretical arguments, and interpretations are the author's own.

https://doi.org/10.33774/apsa-2026-3lm69 **ORCID:** https://orcid.org/0000-0002-2763-2126 Content not peer-reviewed by APSA. **License:** CC BY-NC-ND 4.0