

Fasang, Anette Eva

Article — Published Version

Comment: What's the Added Value?

Sociological Methodology

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Fasang, Anette Eva (2015) : Comment: What's the Added Value?, Sociological Methodology, ISSN 1467-9531, Sage Publications, Thousand Oaks, CA, Vol. 45, Iss. 1, pp. 56–70-, <https://doi.org/10.1177/0081175015587276>

This Version is available at:
<https://hdl.handle.net/10419/190500>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

COMMENT: WHAT'S THE ADDED VALUE?

*Anette Eva Fasang**

*Humboldt University of Berlin and WZB Berlin Social Science Center, Berlin, Germany

Corresponding Author: Anette Eva Fasang, anette.fasang@wzb.eu
DOI: 10.1177/0081175015587276

In the past decade, numerous technical innovations of sequence analysis have been introduced in the social sciences. Recently, scholars have voiced increasing concerns that the development of appropriate theoretical concepts to guide these innovations has lagged behind (Blanchard, Bühlmann, and Gauthier 2014). There is some irony in this, considering that the original motivation for sequence analysis in the social sciences was strongly rooted in theoretical concerns about how we treat process and temporality, as formulated in Andrew Abbott's (1992) notes on narrative positivism. In this comment, I take Robette, Bry, and Lelièvre's (RBL) globally interdependent multiple sequence analysis (GIMSA) (this volume, pp. 1–44) as one of the recent technical propositions, and I argue

(1) for establishing stronger theoretical foundations for new sequence analysis techniques and (2) for more carefully scrutinizing their added value in addition to existing techniques. I first consider possibilities for theorizing processes of global interdependence in linked life-courses more thoroughly and then raise doubts about the added value of GIMSA over the more straightforward technique of multichannel sequence analysis (MCSA) to answer similar questions of regularities in dyadic sequences. These doubts are based on robustness checks conducted on RBL's data that show a high sensitivity of the results to different weighting strategies of the partial least square (PLS) components (step 3 of GIMSA). Furthermore, a systematic comparison of GIMSA and MCSA shows that for the example application, the same results are easily obtainable with MCSA, a procedure that involves fewer steps and fewer potentially consequential decisions by the researcher.

1. THEORETICAL CONSIDERATIONS ON GLOBAL INTERDEPENDENCE IN LINKED LIFE-COURSES

GIMSA is introduced as a new approach for studying intergenerational regularities as “process outcomes” of life-course trajectories rather than “point in time outcomes” such as occupational prestige at age 36 or “trend outcomes” such as the gender wage gap over time (Abbott 2005). Why should we care about global interdependence in process outcomes? To date, intergenerational transmission or inheritance is generally defined as similarity in point-in-time outcomes that exists if parents and their children show the same behavior, usually measured at a specific age. The focus on point-in-time outcomes potentially obscures intergenerational regularities in parent and child behavior, if they are similar in one outcome but not in another that are both part of the same process—for example, if parents and children have the same age of first birth but differ widely in completed fertility. In addition, the focus on similarity is conceptually too narrow and unsatisfactory for three reasons.

First, it neglects global similarity—that is, behavior that is similar in some way but not exactly the same. One example would be if children's family formation is a delayed and protracted version of their parents' family formation. As also noted by RBL, given changing macrostructural contexts from one generation to the next, the same behavior of parents and children might not carry the same implications in different contexts. For instance, giving birth at age 17 was not unusual in the 1960s but is considered early today. Looking for simple one-on-one similarity between parents and their children is of little use, when even

the same outcomes do not necessarily carry the same information about a person's location within a population. Rather, we need to ask, which form can intergenerational inheritance or transmission possibly take given profound macrostructural change from one generation to the next?

Second, a narrow focus on similarity neglects systematic regularities beyond some form of "global" resemblance—that is, patterns of systematic contrast between parents and their children. In critiquing the conventional focus on transmission as similarity, we therefore recently proposed the concept of intergenerational patterns of family formation (Fasang and Raab 2014). Intergenerational patterns denote regularities in family formation trajectories as process outcomes, where specific parents have specific children, but parents and children are not necessarily the same—*global interdependence* in RBL's terminology. Using MCSA, we showed among other groups of transmission a "contrast pattern" of family formation for middle-class American families.

Third, a narrow focus on similarity fails to encompass intergenerational regularities cutting across life domains, such that mothers' employment trajectories might be linked to their daughters' family formation. For instance, daughters of mothers who tried to combine work and family under difficult conditions in the 1960s might choose either work or family because they shy away from a having-it-all model that they observed as difficult to realize. On the basis of the life-course paradigm (Elder, Johnson, and Crosnoe 2003), global regularities that cut across life domains in multidimensional and linked lives are theoretically plausible, but at present we know next to nothing about them.

Developing appropriate methodology for studying such "global" intergenerational regularities in process outcomes beyond simplistic similarity of parents and children in point-in-time outcomes is therefore much needed. GIMSA is motivated by this challenge, and I very much welcome the authors' efforts. However, if GIMSA is to be fruitful in future research, I believe more elaborated theoretical considerations are necessary rather than allusions to the possible general importance of global interdependence, as suggested by RBL. They do not explain why global and not local interdependence is more likely in their example of mothers' and daughters' employment trajectories, nor do they clarify which substantive content of global interdependence is theoretically likely in the study population. This theoretical "aimlessness" is visible in their difficulty to make sense of the rather confusing results of a 10-group typology without a clear takeaway message.

In fairness, developing more careful theoretical arguments about expected globally interdependent regularities is not a flaw of GIMSA as a methodological tool but a collective task for future research. As a starting point, this requires (1) specifying conditions under which global, local, or no interdependence is likely in dyadic sequences and (2) drawing on established sociological theory to specify hypotheses about which substantive content patterns of global interdependence will take, rather than just “fishing for patterns” (Wu 2000).

2. WHAT’S THE ADDED VALUE OF GIMSA?

Despite my enthusiasm for recognizing the importance of global interdependence in process outcomes, I have reservations about the added value of GIMSA. First, I am concerned about the large number of decisions that must be made in the multistep approach. Second, given this complicated implementation, I am not convinced that GIMSA provides sufficient added value over the more straightforward approach of MCSA to study global regularities in dyadic sequences. These concerns are amplified by the chosen example in RBL of mothers’ and daughters’ employment trajectories in France, which seems ill-suited to demonstrate a potential added value of GIMSA.

2.1. *Example Application*

A good empirical example to demonstrate new methods should explicate the added value of the method over other approaches with accessible and straightforward results that could clearly not have been obtained in any other way. Mothers’ and daughters’ employment patterns, however, are precisely not a straightforward case of global interdependence. Arguably two main sources of differences in mothers’ and daughters’ employment trajectories in the second half of the twentieth century in France will be (1) a delay of daughters’ employment onset due to educational expansion and (2) more frequent interruptions in daughters’ employment careers as they strive to combine work and family while entering the labor market in greater numbers than their mothers. These are mainly timing differences of delay and interruption. Issues of timing, however, are a form of local interdependence, as also noted by RBL. Rather than choosing a cross-domain application (e.g., of mothers’ employment and daughters’ family formation) the difference in the example application merely lies in longer observation periods for the

mothers and a slightly different alphabet of employment states, which are both driven by data restrictions, not by substantive consideration. These weaknesses of the example application are apparent in the confusing detailed results that in my opinion struggle to demonstrate an added value of GIMSA, particularly without prior theoretical considerations about which patterns might be expected. However, an unconvincing choice of example is not necessarily a flaw of the method itself, which leads to my core concerns about GIMSA.

2.2. *Complicated Multistep Procedure*

GIMSA proceeds in four steps, each of which requires nontrivial decisions by the researcher. Thorough robustness checks would require one to modify multiple parameters in each decision step: dissimilarity measure to determine sequence similarity, multidimensional scaling (MDS), canonical PLS, and clustering. In practice, it is easy to lose track of what one is actually doing in such a procedure and what might possibly be generating the final results. When replicating this study with the data and code provided by RBL, I found that the results were highly sensitive to different strategies for weighting the PLS components before generating the distance matrix for the clustering. Weighting the PLS components, RBL argue, is important to ensure that the results are not merely driven by different degrees of heterogeneity in the sequences in the two dimensions (here mothers and daughters). RBL conduct several robustness checks, keeping the number of clusters fixed at 10 groups for each weighting approach. They report considerably stable typologies—when the number of clusters is fixed at 10. However, under different weighting scenarios, cluster cutoff criteria support very different cluster solutions that do not justify retaining 10 groups for each scenario—panels A to D of Figure 1. In practice, the applied user, whose choice of clusters is guided by cluster cutoff criteria, would arrive at different numbers of clusters with vastly different substantive conclusions, depending on which weighting of the PLS component was chosen. Figure 1 visualizes cluster cutoff criteria (Studer 2013) for different weighting strategies of the PLS components. Panel A on the left side in Figures 2, 3, and 4 shows three cluster solutions supported by these cutoff criteria after GIMSA. Panel B on the right side in Figures 2, 3, and 4 compares them with results obtained with different specifications of MCSA that are discussed below.

Specifically, in scenario 1 without weighting the PLS components, cluster cutoff criteria in panel A of Figure 1 support a 4-cluster solution,

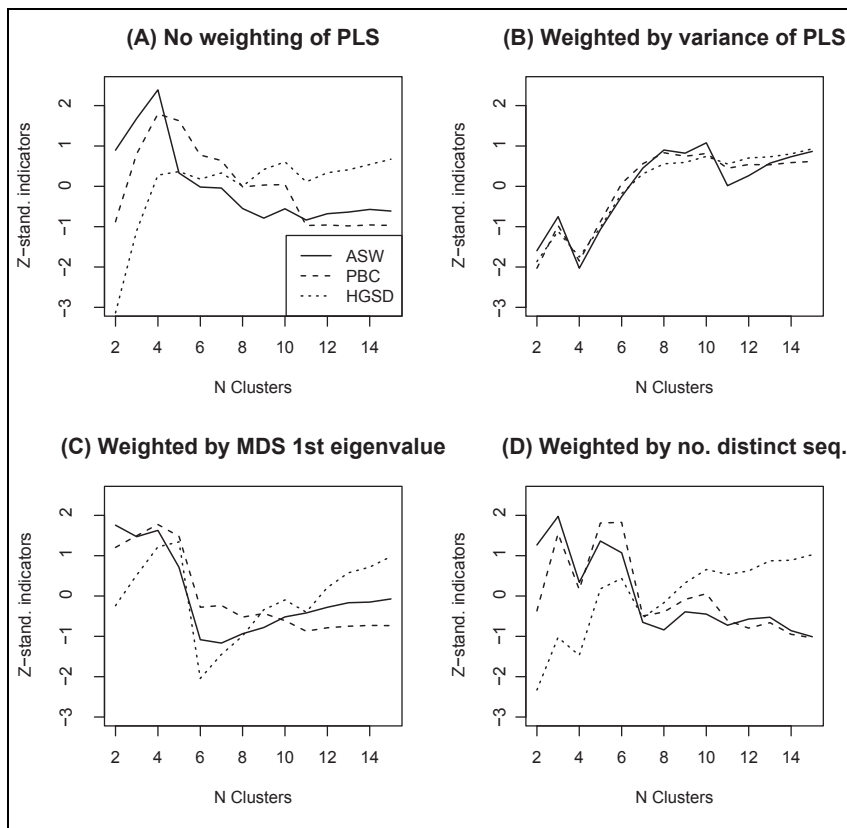


Figure 1. Cluster cutoff criteria by four weighting scenarios of PLS in RBL.

Note: ASW = average silhouette width; PBC = point biserial correlation; PLS = partial least square; HGSD = Hubert’s gamma (Sommer’s *D*); RBL = Robette, Bry, and Lelièvre.

in which all groups of daughters look very similar. The mothers are divided into 4 groups depending on the one state that dominated their lifetime employment (clusters in Figure 2A). We can conclude that there are 4 distinct groups of mothers (average silhouette width [ASW] = .42), but there are no systematic intergenerational links between mother and daughter trajectories. Scenario B as reported in RBL, weighting by the variance of the PLS components, yields 10 clusters as the best solution, supported by a much lower ASW of .20, meaning that this is a much less discriminant grouping (Figure 1B). However, we might conclude that there is some complex global interdependence between the mother and daughter trajectories as reported in RBL.

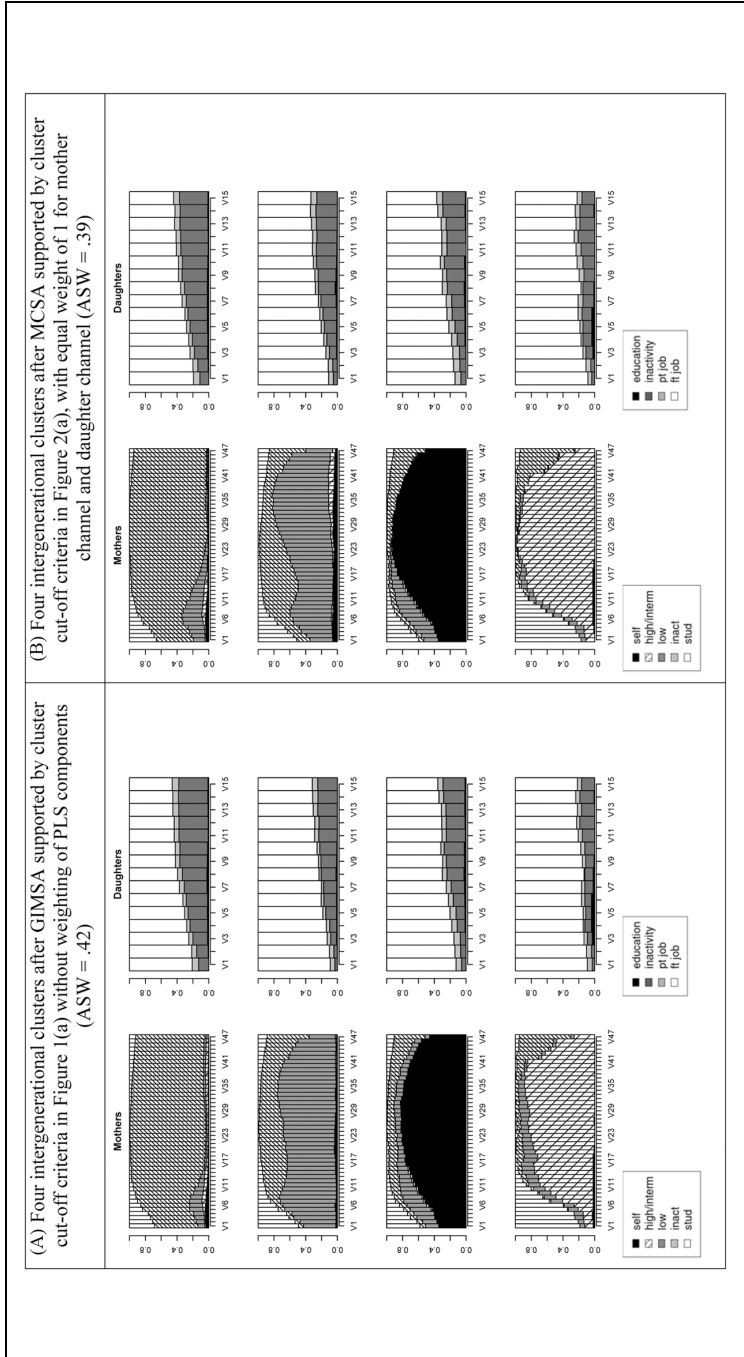


Figure 2. Comparison of clusters obtained with GIMSA without weighting of PLS and MCSA with equal weights of 1 for mothers and daughters. (A) Four intergenerational clusters after GIMSA supported by cluster cutoff criteria in Figure 1A without weighting of PLS components (ASW = .42). (B) Four intergenerational clusters after MCSA supported by cluster cutoff criteria in Figure 5A, with equal weight of 1 for mother channel and daughter channel (ASW = .39).
Note: ASW = average silhouette width; ft = full-time; GIMSA = globally interdependent multiple sequence analysis; inact = inactivity; interm = intermediate; MCSA = multichannel sequence analysis; PLS = partial least square; pt = part-time; stud = student.

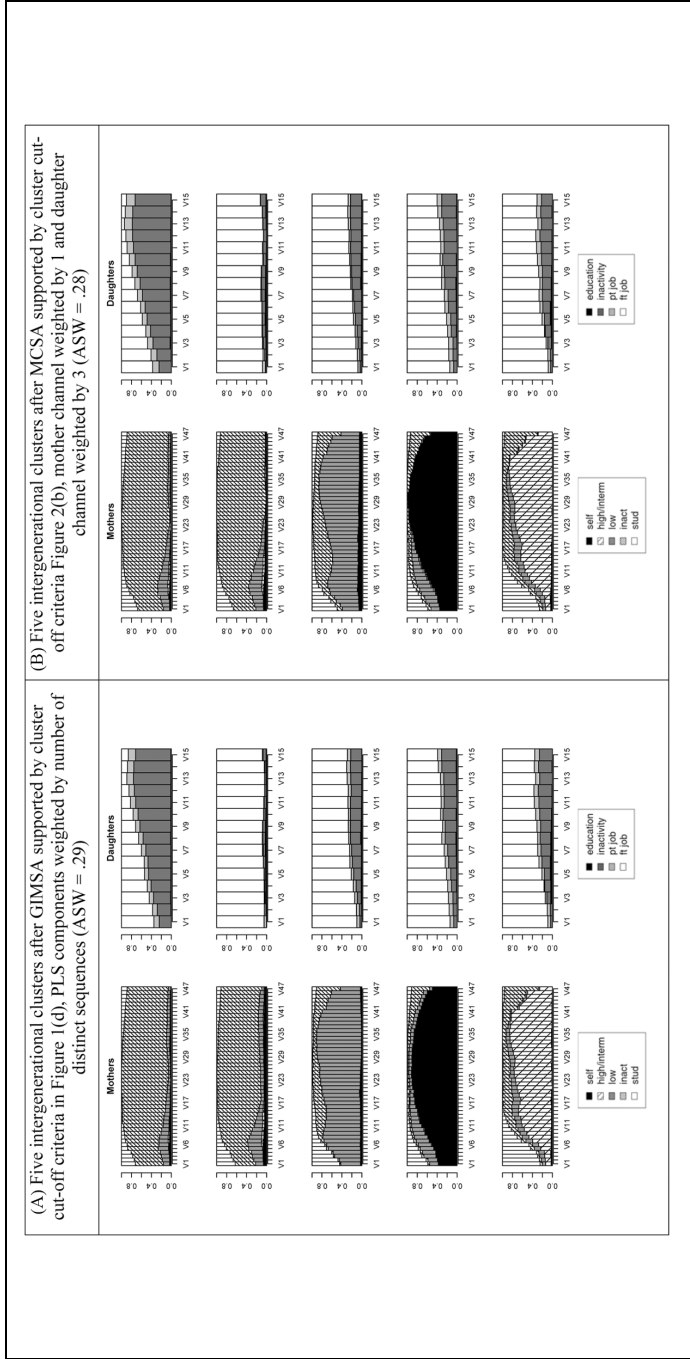


Figure 3. Comparison of clusters obtained with GIMSA with PLS weighted by number of distinct sequences and MCSA with daughter channel weighted three times the mother channel. (A) Five intergenerational clusters after GIMSA supported by cluster cutoff criteria in Figure 1D, PLS components weighted by number of distinct sequences (ASW = .29). (B) Five intergenerational clusters after MCSA supported by cluster cutoff criteria Figure 5B, mother channel weighted by 1 and daughter channel weighted by 3 (ASW = .28).

Note: ASW = average silhouette width; ft = full-time; GIMSA = globally interdependent multiple sequence analysis; inact = inactivity; interm = intermediate; MCSA = multichannel sequence analysis; PLS = partial least square; pt = part-time; stud = student.

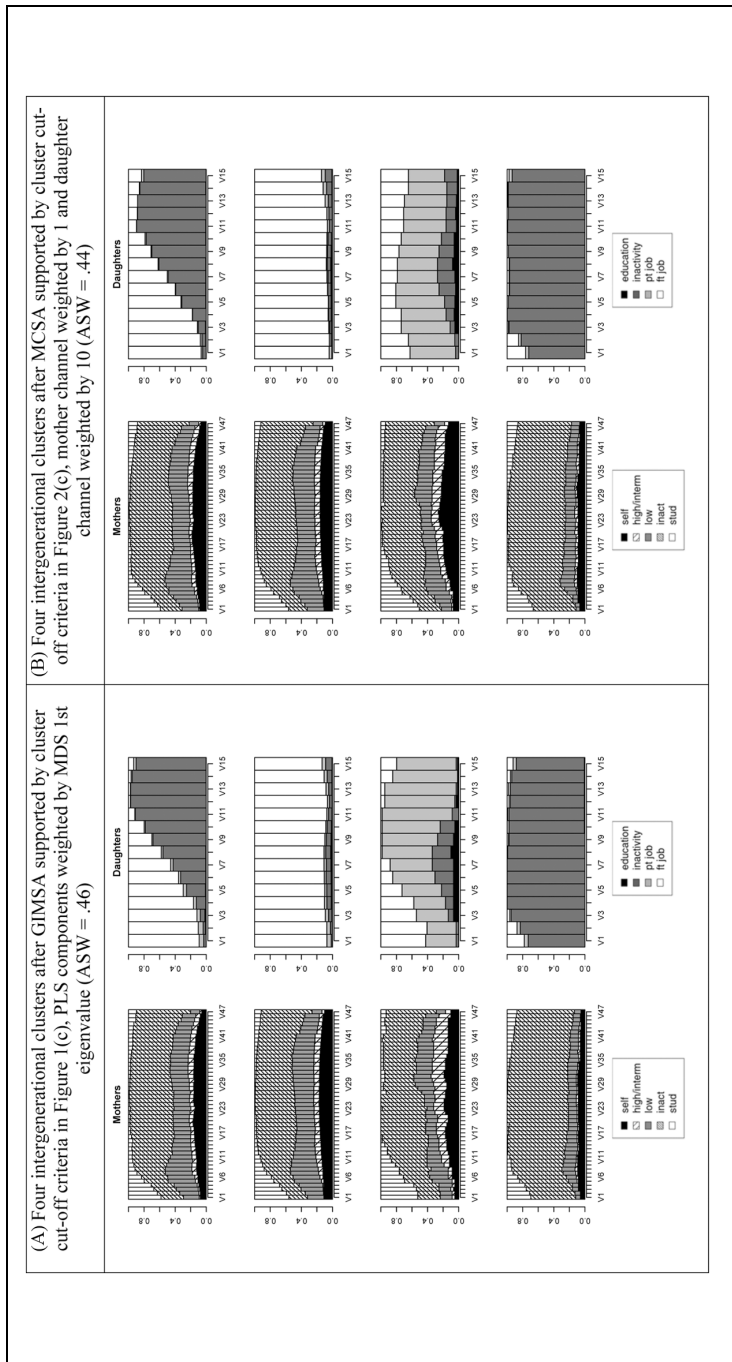


Figure 4. Comparison of clusters obtained with GIMSA with PLS weighted by MDS first eigenvalue and MCSA with daughter channel weighted 10 times the mother channel. (A) Four intergenerational clusters after GIMSA supported by cluster cutoff criteria in Figure 1C, PLS components weighted by MDS first eigenvalue (ASW = .46). (B) Four intergenerational clusters after MCSA supported by cluster cutoff criteria in Figure 5C, mother channel weighted by 1 and daughter channel weighted by 10 (ASW = .44).

Note: ASW = average silhouette width; ft = full-time; GIMSA = globally interdependent multiple sequence analysis; inact = inactivity; interm = intermediate; MCSA = multichannel sequence analysis; PLS = partial least square; pt = part-time; stud = student.

Weighting by the MDS first eigenvalue in scenario C yields either 2 clusters at an ASW of .47 or 4 clusters at an ASW of .46, which again is a much more discriminant grouping than for the originally proposed 10 clusters (Figure 1C). The 4 clusters in this case are strongly patterned by the dominant state of the daughters (Figure 4A): one group that goes from full-time to inactivity, one working full-time, one working part-time, and a predominantly inactive group. The 4 mother clusters are very similar. Again we would conclude that there is no intergenerational interdependence in the mother and daughter sequences (Figure 4A). In scenario D, when weighting by the number of distinct sequences, either a 3-cluster (ASW = .31) or a 5-cluster (ASW = .29) solution is supported by the cutoff criteria (Figure 1D). The 5-cluster solution shows two groups, in which the mother is inactive, one in which daughters move from full-time work to inactivity and one in which daughters work full-time throughout (Figure 3A). This does suggest some global interdependence in intergenerational employment patterns for mother-daughter dyads in which the mother was mostly inactive.

In sum, two of these weighting scenarios of the PLS components support no global interdependence, while two support different substantive contents of global interdependence. What am I to conclude from these different results, derived by varying just one among many choices necessary in GIMSA? This is particularly so because I lack any concrete theoretical rationale about which extent and substantive content of global interdependence I might be looking for. These results may simply reflect the ill-suited example application, if in this case there is no interdependence, or if local and not global interdependence is the dominant link between mothers' and daughters' employment. Possibly, in an example application in which strong global interdependence exists, GIMSA would consistently find it. The foregoing example, however, does draw attention to the potential pitfalls of methodological procedures involving many steps with many decisions to be made. This in turn warrants careful consideration of whether we can obtain the same results in a more simple and straightforward way.

2.3. *Added Value to MCSA*

Next to cross-tabulating separate cluster solutions for the mother and daughter sequences (see RBL), MCSA is the most obvious alternative to uncover patterns of global interdependence (Fasang and Raab 2014;

Gauthier et al. 2010; Pollock 2007). In MCSA, two members of a dyad—for example, a mother and daughter—are treated as two separate channels of an intergenerational dyadic sequence. MCSA is more straightforward to apply and can identify patterns of global interdependence in fewer analysis steps, and thus fewer possibly consequential decisions are necessary from the researcher. The researcher specifies channel-specific substitution costs and weights attached to each channel and executes the MCSA that generates a pairwise distance matrix, which is entered into a cluster analysis. Compared with GIMSA, all decisions related to the MDS and PLS are omitted. RBL argue that MCSA is not practical for uncovering patterns of global interdependence because it requires sequences of equal length on both dimensions. A simple and straightforward way to deal with sequences of unequal length is to insert a missing value state to fill the gap between the shorter and longer sequence dimensions.

After filling in the remaining time of the daughter sequences with a missing value state, I used MCSA on mothers' and daughters' employment histories with constant substitution costs of 2 and indel costs of 1 as a standard choice (MacIndoe and Abbott 2004). Depending on how the mother and the daughter sequences were weighted, I found substantively the same results as in different weighting scenarios of the PLS in GIMSA (Figures 2–4). Three weighting scenarios that place different emphasis on the longer and shorter sequence dimension were tested: equal weights of one for the mother and daughter channel, weighting the daughter channel 3 times the mother channel, and weighting the daughter channel 10 times the mother channel (see cluster cutoff criteria in Figure 5). Figures 2 to 4 show systematic comparisons of typologies supported by the cluster cutoff criteria in Figures 1 and 5 obtained with GIMSA and MCSA.

Figure 2 shows a four-group typology supported for GIMSA without weights of the PLS components (Figure 1A) and MCSA with equal weights of 1 for the mother and daughter channel. The correlation between the two cluster solutions is .86, and the clusters in the two panels would lead to the same substantive conclusion: there is strong patterning into four groups in the mother generation and no intergenerational global interdependence.

This might result from the longer observation periods and correspondingly higher potential variation in the mother sequences that are roughly three times as long as the daughter sequences. To account for this different sequence length and the resulting greater possible variation

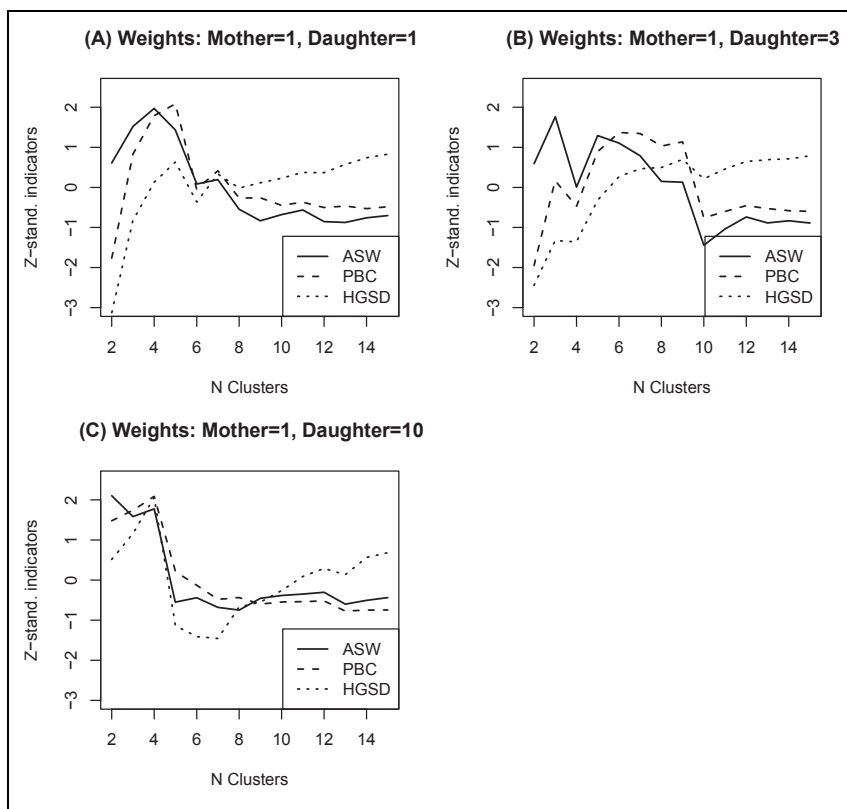


Figure 5. Cluster cutoff criteria after MCSA with missing value states in daughter sequences in different weighting scenarios of the two channels.
 Note: ASW = average silhouette width; PBC = point biserial correlation; HGSD = Hubert’s gamma (Sommer’s *D*); MCSA = multichannel sequence analysis.

for mothers, I repeated MCSA, weighting the daughter channel three times the mother channel. Figure 3 compares the five-cluster solution supported by the cutoff criteria for this MCSA with GIMSA weighting the PLS by the number of distinct sequences. The results support some global intergenerational dependence, such that dyads in which mothers were mainly inactive are split into a group in which the daughters were consistently employed full-time and a group in which the daughters were employed full-time but then became inactive. Again, in this comparison, GIMSA and MCSA would lead to the exact same substantive conclusions.

Figure 4 compares GIMSA weighted by the MDS first eigenvalue and MCSA when weighting the daughter sequences 10 times the mother

sequences. I find four identical clusters supported by the cutoff criteria in Figures 1 and 5 that are driven mainly by differences in the daughter sequences. Finally, allowing for 10 clusters after MCSA weighting the daughter sequences three times the mother sequences shows 10 groups (not presented here) that correspond closely to what is found with GIMSA and presented by RBL. Note that these 10 groups are, however, not supported by the cluster cutoff criteria after MCSA, although there is some support for a 9-cluster solution (Figure 5B).

The bottom line is that in this application, MCSA with missing value states for the shorter sequences yields the same substantive results as GIMSA, while requiring fewer necessary decisions from the researcher. Therefore, MCSA is more applicable and transparent without any apparent disadvantage compared with GIMSA in this example application. Attaching simple numerical weights to the two channels in MCSA is more intuitive than the complex weighting of the PLS components by MDS eigenvalues, variances, and the number of distinct sequences. Weighting of the channels also allows researchers to account for unequal sequence length by placing higher weights on the shorter channel filled with a missing value state. It is straightforward to include different alphabets of states in the two channels in MCSA. Because transformation costs can be defined separately for each channel, different rationales for sequence similarity in the two dimensions are easy to implement. When gaps are filled with missing value states, it is also possible to consider dimensions with different time units. Technically, MCSA easily extends to triads and beyond if a researcher wants to examine multigenerational patterns by including additional channels. In contrast to RBL, I therefore find the simple procedure of filling in gaps with missing value states in MCSA more practical—not less—and less error prone compared with the four-step procedure of GIMSA.

I want to conclude by emphasizing my enthusiasm for the attention given to improving the methods for analyzing patterns of global interdependence in linked life-course sequences. Furthermore, I do not want to jump to conclusions. There might be sociologically important and interesting empirical cases in which GIMSA has added value over MCSA. For instance, with different time units and extremely unequal sequence length, GIMSA might indeed be more useful. For the reasons outlined above, on the basis of the example of mothers' and daughters' employment trajectories in France, I am not (yet) convinced that GIMSA provides sufficient added value to cross-tabulations of separate cluster solutions or MCSA to make the complicated multistep procedure worthwhile.

To push the analysis of dyadic sequences forward, a useful step in future research would certainly be a simple test to decide whether there are associations—any dependence—in two sequence dimensions at all before continuing to disentangle whether these are local or global. More generally, if sequence analysis is to do more for us, we need to push rigorous theorizing and to gather more empirical applications, including systematic replications and reproductions of previous work. One important task for researchers is to delineate the complementary added value of different sequence analysis techniques for specific research questions among themselves as well as in comparison with adjacent methods, such as group-based trajectory modeling or latent class models. Methods are tools for meeting ends, but they are not ends in themselves. We should therefore judge them by their applicability and usefulness in helping us address sociologically important questions.

Acknowledgments

I thank Marcel Raab for fruitful discussions related to this commentary and Zachary Van Winkle for editing and proofreading.

References

- Abbott, Andrew. 1992. "From Causes to Events: Notes on Narrative Positivism." *Sociological Methods and Research* 20(4):428–55.
- Abbott, Andrew. 2005. "The Idea of Outcome in U.S. Sociology." Pp. 393–426 in *The Politics of Method in the Human Sciences*, edited by G. Steinmetz. Durham, NC: Duke University Press.
- Blanchard, Philippe, Felix Bühlmann, and Jacques-Antoine Gauthier, eds. 2014. *Advances in Sequence Analysis: Theory, Method, Applications*. New York: Springer.
- Elder, Glen H., Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. "The Emergence and Development of Life Course Theory." Pp. 3–19 in *Handbook of the Life Course*, edited by J. T. Mortimer and M. J. Shanahan. New York: Kluwer Academic/Plenum.
- Fasang, Anette E., and Marcel Raab. 2014. "Beyond Transmission: Intergenerational Patterns of Family Formation among Middle-class American Families." *Demography* 51(5):1703–28.
- Gauthier, Jacques-Antoine, Eric D. Widmer, Philipp Bucher, and Cédric Notredame. 2010. "Multichannel Sequence Analysis Applied to Social Science Data." Pp. 1–38 in *Sociological Methodology*, vol. 40, edited by Tim Futing Liao. Hoboken, NJ: Wiley-Blackwell.
- MacIndoe, Heather, and Andrew Abbott. 2004. "Sequence Analysis and Optimal Matching Techniques for Social Science Data." Pp. 387–406 in *Handbook of Data Analysis*, edited by A. Bryman and M. Hardy. London: Sage Ltd.
- Pollock, Gary. 2007. "Holistic Trajectories: A Study of Combined Employment, Housing, and Family Careers by Using Multiple-sequence Analysis." *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 170(1), 167–83.

- Studer, Matthias. 2013. "Weighted Cluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R." *LIVES Working Papers*, 24.
- Wu, Lawrence L. 2000. "Some Comments on 'Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect.'" *Sociological Methods and Research* 29(1):41–64.

Author Biography

Anette Eva Fasang is a professor of microsociology at Humboldt University of Berlin and head of the demography and inequality research group at the WZB Berlin Social Science Center. She obtained her PhD from Jacobs University Bremen and completed postdoctoral research at Yale University and Columbia University. Her research interests include social demography, stratification, life-course sociology, family demography, and methods for longitudinal data analysis.
