CAN YOU AGGREGATE INDIVIDUAL-LEVEL DATA?

Based on: Van Mierlo, H., Vermunt, J., & Rutte, C. (2009). Composing Group-level Constructs From Individual-level Survey Data. *Organizational Research Methods*, 12(2), 368-392.

Data is often aggregated without much explication of the aggregation method, which can result in several problems and make the interpretation of results difficult. Furthermore, data aggregation often happens for pragmatic reasons, as it can be difficult to collect data at the exact level desired. However, it is possible to examine data to better understand whether aggregation is appropriate in a specific case.

There are several ways to compose group-level constructs. Two of the most widely used methods in organizational settings are *direct-consensus composition* and *referent-shift consensus composition*.

- **Direct-consensus composition**: This method is based on individual-level consensus, so items on the group level refer to individual perceptions.
    - Example: *Do your work activities require creativity?*
- **Referent-shift consensus composition**: This method is based on individual responses but is modified to fit a group setting. Therefore, items refer to the group.
    - Example: *Is creativity required in your team's work activities?*

Examining the differences between composition methods

In the following, we briefly summarize some ways to determine which composition type is appropriate.

1. Similarity of constructs: A <u>factor analysis</u> between groups can help uncover whether certain items load on distinct factors. This can indicate whether the two composition types yield distinct group constructs (see Figure 1).

2. Similarity of constructs: <u>Correlations</u> can provide an indication of the possible overlap between the two composition types. Furthermore, correlation within groups can indicate whether respondents distinguish between their own situation and that of the group. Correlations between groups show the extent to which scales are distinct **after** aggregating to the group level.
    - Example: Low shared variance between groups indicates that the two composition types are different. In contrast, high shared variance between groups indicates an overlap between composition types.
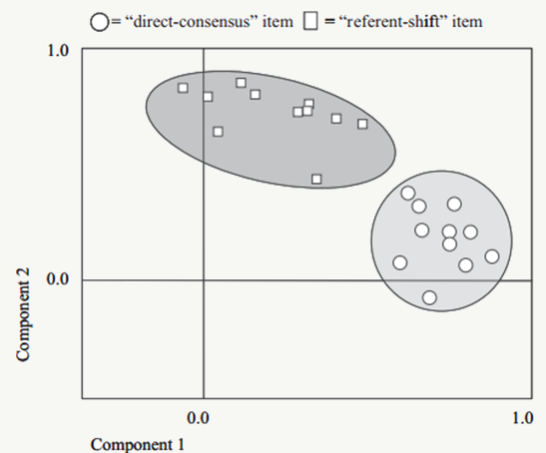

Figure 1. Graphical illustration of **factor analysis**. The two composition types load on different components, which indicates that they yield distinct constructs. Source: Van Mierlo, Vermunt and Rutte (2009).

- Example: Low shared variance between items at the individual level and the within-group level indicates that members of the group differentiate between their own beliefs and those of the group.

3. Reliability: <u>Variance</u> within and between groups can help investigate whether the two composition methods yield reliable constructs at the group level. A group-level construct should be (relatively) homogeneous within groups and heterogeneous between groups to be reliable.
    - Example: If all of the groups have the same scores, there is no difference between the groups. Therefore, the construct is not very reliable. Conversely, if the between-group variance results in different scores, then there is a difference between groups.

4. Construct validity: <u>Agreement</u> among group members (i.e., "interrater agreement") can indicate the extent of consistency in the responses of group members. One way to investigate this is to compare the observed variance in the group for some of the items to the expected variance if members responded randomly (for more on technical measurement, see Lindell, Brandt and Whitney,[1] or Dunlap, Burke and Smith-Crowe[2]). High variance could mean that members' individual responses about their own perceptions cannot be aggregated to the group level.
    - Example: If the difference between the items among members of the same group is large, this could indicate lower observed agreement than what would be expected in random conditions.

5. Construct validity: A within-group <u>factor analysis</u> can help investigate whether the group-level construct actually captures a group-level phenomenon. In *direct-consensus composition*, the items are based on the individual. In *referent-shift consensus*, the items are based on the group and not meant to measure individual differences.
    - Example: If a factor analysis for direct consensus yields a one-factor structure, then it captures individual differences both within and between groups.
    - Example: If a factor analysis for referent-shift consensus shows no meaningful structure, then all members answered in a similar way.

These are just a few of the ways in which one can examine survey data. They provide a better understanding of how individuals respond to and perceive a questionnaire and its items. This can help you understand whether you can aggregate your data for further analysis and how you should do so.

---

[1] Lindell, M. K., Brandt, C. J., and Whitney, D. J. (1999). 'A Revised Index of Interrater Agreement for Multi-item ratings of a single target'. *Applied Psychological Measurement*. 23:127-135.
[2] Dunlap, W. P., Burke, M. J., and Smith-Crowe, K. (2003). 'Accurate Tests of Statistical Significance for rWG and Average Deviation Interrater Agreement Indexes'. *Journal of Applied Psychology*, 88: 356-362.