

Middle East Technical University

Department of Computer Engineering

CENG 574 — Statistical Data Analysis

Assignment 1

Spring 2025–2026

Kerem Recep Gür | 2448462

1. Summary of the Article

The article by Li et al. (2022), titled “*Evaluating the Impact of Big Data Analytics Usage on the Decision-Making Quality of Organizations*”, investigates how the extent to which firms use big data analytics tools influences the quality of their organizational decisions. The study is grounded in **dynamic capability theory**, which distinguishes between low-order and high-order organizational capabilities: big data analytics usage is treated as a low-order capability, while data analytics capabilities (encompassing technical skills and managerial skills) are treated as a high-order capability that can be cultivated through lower-order practices.

The central argument of the paper is that big data analytics usage does not improve decision-making quality in isolation; rather, its full benefit is realized when it develops the firm’s underlying data analytics capabilities. Accordingly, the authors propose and test three hypotheses:

- **H1:** Big data analytics usage positively affects decision-making quality.
- **H2:** Big data analytics usage positively affects data analytics capabilities.
- **H3:** Data analytics capabilities positively affect decision-making quality.

Data were collected via a structured questionnaire distributed to senior executives of provincial agricultural firms in China, with the assistance of China’s Department of Agriculture and Rural Affairs. Out of 2,652 firms contacted, 286 responses were received, of which 240 were valid after excluding incomplete answers. All measurement items were adapted from validated scales in prior information-systems literature and translated into Chinese by a bilingual co-author.

The empirical analysis was carried out using **Partial Least Squares Structural Equation Modeling (PLS-SEM)** via SmartPLS 3.3.2. All three hypotheses were

supported: big data analytics usage had a direct positive effect on decision-making quality ($\beta = 0.223$, $p < 0.01$) and a strong positive effect on data analytics capabilities ($\beta = 0.623$, $p < 0.001$), while data analytics capabilities in turn positively influenced decision-making quality ($\beta = 0.394$, $p < 0.001$). Mediation analysis using both the Sobel test and bootstrapping confirmed that data analytics capabilities *partially* mediate the relationship between big data analytics usage and decision-making quality (indirect effect = 0.218, 95% CI: [0.123, 0.329]). Robustness checks, including an alternative model reversal and confirmatory composite analysis, further validated the proposed model.

The authors conclude that firms should not only promote the adoption of big data analytics tools across their operations but also deliberately invest in developing their employees' technical and managerial skills, since these capabilities serve as the key mechanism through which analytics usage ultimately translates into higher-quality decisions.

2. Strengths of the Study

2.1. Solid Theoretical Grounding

The study is built on a well-established theoretical lens—dynamic capability theory—which provides a principled rationale for why big data analytics usage (a lower-order capability) should cultivate data analytics capabilities (a higher-order capability) and thereby improve decision-making quality. Anchoring the research model in theory prevents it from being merely descriptive and allows its findings to generalize beyond the agricultural context.

2.2. Rigorous Scale Development and Validation

All measurement items were adapted from previously validated scales in the IS literature (Yunis et al., 2018; Ghasemaghaei, 2019; Ghasemaghaei et al., 2018). The authors subjected the scales to a multi-step validation process: forward–backward translation, two independent bilingual proofreaders, and a pretest with 20 students majoring in agricultural big data analysis. Convergent validity ($AVE \geq 0.510$), composite reliability ($CR \geq 0.822$), and internal consistency (Cronbach's $\alpha \geq 0.774$) all exceeded recommended thresholds.

2.3. Addressing Common Method Bias

Because both independent and dependent variables were collected from the same respondents in the same survey, common method bias is a legitimate concern. The

authors proactively applied the *marker variable technique* (Lindell & Whitney, 2001), using gender as a theoretically unrelated marker. The negligible average correlation ($r_M = 0.014$) and the statistical equivalence of path estimates between the baseline and adjusted models ($\chi^2 = 9$, $p = 0.109$) indicate that common method bias did not substantially distort the results.

2.4. Non-Response Bias Assessment

The authors compared early ($n = 185$) and late ($n = 55$) respondents on all key constructs using t-tests and chi-square tests. The absence of significant differences between the groups strengthens confidence that the achieved sample is representative of the target population.

2.5. Comprehensive Robustness Testing

Beyond the main PLS analysis, the authors re-ran the model treating all constructs as composite indicators, performed confirmatory composite analysis (SRMR = 0.034, below the 0.08 threshold), and estimated an alternative model in which the direction of the BDU–DAC relationship was reversed. The alternative model produced inferior fit statistics, lending additional support to the theoretical direction of effects.

2.6. Dual Mediation Test

Using both the Sobel test and bias-corrected bootstrapping ($n = 4,999$ resamples) to assess mediation is considered best practice in PLS-SEM research, as bootstrapping does not assume normality of the sampling distribution. The consistent significance across both methods strengthens the mediation conclusion.

3. Weaknesses of the Study and Suggested Improvements

3.1. Single-Sector, Single-Country Sample

The study relies exclusively on agricultural firms in China, which raises questions about external validity. The big data landscape, regulatory environment, and organizational culture in Chinese agriculture may differ considerably from other industries (e.g., manufacturing, finance, healthcare) or other national contexts. As a result, it is unclear whether the reported effect sizes would hold in different settings.

What could be done: Future research should replicate the model across multiple industries and countries. A cross-industry design would also allow moderation analysis to test whether industry type strengthens or weakens the relationship between

big data analytics usage and decision-making quality.

3.2. Cross-Sectional Research Design

All data were collected at a single point in time, meaning that the study cannot establish causal ordering or capture the temporal dynamics of capability building. Firms may need months or years of sustained analytics usage before observable improvements in data analytics capabilities and decision quality materialize.

What could be done: A longitudinal panel design—measuring the same firms at multiple time points—would allow researchers to examine how capabilities evolve over time and whether the effects persist, strengthen, or decay. Alternatively, a quasi-experimental or difference-in-differences design exploiting exogenous variation in analytics adoption (e.g., regulatory mandates or subsidized technology roll-outs) could support stronger causal inference.

3.3. Low Explained Variance in Decision-Making Quality

The model accounts for only 32.5% of the variance in decision-making quality ($R^2 = 0.325$). While this is acceptable in exploratory IS research, it implies that the majority of variation in decision-making quality is driven by factors outside the model—such as leadership quality, organizational learning culture, data governance, or market uncertainty.

What could be done: The model should be extended with theoretically motivated moderators and additional mediators. For example, *organizational absorptive capacity*, *data governance maturity*, and *top-management support* have been identified in adjacent literature as important antecedents of big data value. Including such variables would both raise explanatory power and provide richer managerial guidance.

3.4. Discriminant Validity Concern for BDU and DAC

The HTMT ratio between big data analytics usage and data analytics capabilities is 0.857, which is above the stricter recommended threshold of 0.850 (though below the more lenient threshold of 0.900). This suggests that respondents may not have sharply distinguished between *using* analytics tools and *being capable* with analytics, which is conceptually important for the study’s mediation argument.

What could be done: The measurement model could be refined by adding items that more concretely differentiate operational usage (frequency, breadth of tool deployment) from capability (depth of analytical expertise, ability to interpret and act on

results). Qualitative pretesting with industry practitioners could help identify item wordings that produce clearer conceptual separation.

3.5. Perceptual, Self-Reported Measures

All constructs—including decision-making quality—are measured through managerial self-reports on a 5-point Likert scale. Self-reported decision quality is inherently subjective and may be inflated by social desirability or optimism bias. Executives who champion big data initiatives may unconsciously rate their firm’s decisions more favorably.

What could be done: Archival or objective performance indicators (e.g., forecast accuracy, product defect rates, inventory turnover) could be used as observable proxies for decision-making quality alongside self-reported measures. Using both objective and perceptual indicators in a multi-method design would substantially strengthen the construct validity of the dependent variable.

3.6. Absence of Boundary Conditions

The model treats the relationship between big data analytics usage and decision-making quality as uniformly positive across all firms in the sample. However, the IT productivity paradox—briefly acknowledged in the introduction—suggests that returns to analytics usage may depend on firm-level contingencies such as size, IT maturity, or data quality. None of these potential moderators is tested.

What could be done: Including interaction terms between big data analytics usage and candidate moderators (e.g., data quality, IT infrastructure, or competitive intensity) would identify the conditions under which analytics usage yields the greatest improvements in decision quality. This would make the prescriptive implications of the study more precise and actionable for managers.

References

- Li, L., Lin, J., Ouyang, Y., & Luo, X. R. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. *Technological Forecasting & Social Change*, 175, 121355. <https://doi.org/10.1016/j.techfore.2021.121355>