Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

EVALUATING COMPREHENSION OF AGENT-BASED SOCIAL SIMULATION VISUALIZATION TECHNIQUES: A FRAMEWORK BASED ON STATISTICAL LITERACY AND COGNITIVE PROCESSING

Kotaro Ohori[1], Kyoko Kageura[1], and Shohei Yamane[2]

[1]Faculty of Business Administration, Toyo University, Tokyo, JAPAN
[2] Converging Technology Laboratory, Fujitsu, JAPAN

ABSTRACT

Agent-based social simulation (ABSS) has gained attention as a powerful method for analyzing complex social phenomena. However, the visualization of ABSS outputs is often difficult to interpret for users without expertise in ABSS modeling. This study analyzes how statistical literacy affects the comprehension of ABSS visualizations, based on cognitive processes defined in educational psychology. A web-based survey using five typical visualizations based on Schelling's segregation model was conducted in Japan. The results showed a moderate positive correlation between statistical literacy and visualization comprehension, while some visualizations remained difficult to interpret even for participants with high literacy. Further machine learning analysis revealed that model performance varied by cognitive stage, and that basic and applied statistical skills had different impacts on comprehension across stages. These findings provide a foundation for designing visualizations tailored to user characteristics and offer insights for effective communication based on ABSS.

1 INTRODUCTION

Agent-Based Social Simulation (ABSS) has emerged as a powerful method for analyzing complex, multi-factor social phenomena. By modeling the micro-level behaviors of agents, such as individuals or organizations, and observing their interactions, ABSS reveals emergent macro-level patterns (North and Macal, 2007; Gilbert, 2008). Recent developments have combined ABSS with statistical and machine learning techniques to deepen such analysis (Lee et al., 2015; Tanaka et al., 2018; Yamada et al., 2020). However, the outputs generated through such advanced analysis are often multidimensional and voluminous, imposing a high cognitive burden on practitioners or stakeholders who aim to utilize ABSS for decision-making (hereafter, "ABSS users"). Effective decision-making requires the ability to accurately interpret and extract insights from these outputs, making appropriate visualization essential.

In the ABSS domain, various visualization techniques, such as Landscape Analysis of Possibilities (Goto and Takahashi, 2011) and Cladogram Analysis of Possibilities (Goto, 2020), have been proposed to structurally organize multidimensional data and support the comparison between scenarios as well as the visualization of causal structures. However, these methods are often visually and structurally complex, making them difficult to interpret for ABSS users with low levels of statistical or data literacy. In other words, regardless of the sophistication of analysis, if outputs cannot be communicated in an understandable manner, the societal utilization of ABSS will remain limited.

In contrast to the ABSS field, the broader domain of data visualization has actively examined the effects of visualization techniques on human comprehension, particularly from the perspective of non-expert users. For example, Schonlau and Peters (2008) experimentally investigated how the use of tables, bar charts, and pie charts influences accuracy in information interpretation. Padilla et al. (2015) demonstrated that the way uncertainty is visualized, using spatial encodings such as length or area versus non-spatial encodings like color or brightness, can significantly affect decision-making tendencies. Maltese et al. (2015) showed that

979-8-3315-8726-0/25/$31.00 ©2025 IEEE

learners with limited scientific experience often misunderstand how to read or construct graphs, and that accuracy improves with increased domain expertise.

These studies suggest that the visualization formats greatly influence user comprehension, and that user literacy also plays a crucial role. In the field of ABSS, few studies have quantitatively evaluated how different visualization techniques impact user comprehension, as most existing methods are designed from the perspective of expert modelers or analysts, with limited consideration for the understanding of ABSS users. This insufficient consideration of user comprehension may be one reason why ABSS outputs have not yet been widely utilized in real-world decision-making or policy development.

The purpose of this study is to develop a structured analytical framework for evaluating the comprehension of ABSS visualizations, considering both users' levels of statistical literacy and stages of cognitive processing. This framework was applied in an empirical study conducted in Japan as an initial attempt to identify which visualization types facilitate comprehension across user groups with different levels of statistical literacy. We assume that the cognitive process of interpreting ABSS outputs involves multiple stages: (1) identifying data points, (2) interpreting their meaning, (3) comparing changes caused by different simulation settings, and (4) analyzing the underlying mechanisms. To evaluate comprehension at each of these stages in relation to statistical literacy, we applied Bloom's taxonomy of cognitive processes (Bloom, 1956), widely used in educational psychology, to organize and assess visualization comprehension.

The structure of this paper is as follows. Section 2 introduces the proposed analytical framework and describes the design of the empirical study, including the methods used to measure statistical literacy and visualization comprehension. Section 3 presents descriptive and correlational analyses to examine overall trends between statistical literacy and comprehension scores. Section 4 explores, in greater depth, which statistical literacy components most significantly impact comprehension through machine learning-based analysis. Section 5 discusses practical implications for the design of ABSS visualizations based on the findings. Finally, Section 6 summarizes the study's conclusions and suggests future directions.

2 ANALYTICAL FRAMEWORK AND EMPIRICAL STUDY DESIGN

To address the research objective outlined in Section 1, this study first proposes an analytical framework that integrates statistical literacy, cognitive stages, and visualization techniques to evaluate users' comprehension of ABSS outputs. The framework assesses user comprehension across four cognitive stages, Remember, Understand, Apply, and Analyze, adapted from Bloom's taxonomy (Bloom, 1956; Anderson and Krathwohl, 2001). Prior research in data visualization (Burns et al., 2020) has highlighted the value of incorporating cognitive process stages for nuanced analysis of how visualization formats influence user understanding. As shown in Figure 1, comprehension is measured using task items mapped to each combination of cognitive stage and visualization technique. Each cell represents one assessment item targeting a specific combination. Statistical literacy is evaluated using Watson's (1997) three-tiered model. By combining cognitive stages, visualization types, and statistical literacy levels, the framework enables a systematic evaluation of how user comprehension varies across these dimensions. The following sections detail the empirical study built on this framework, including the design of the statistical literacy measurement (Section 2.1) and the visualization comprehension measurement (Section 2.2).

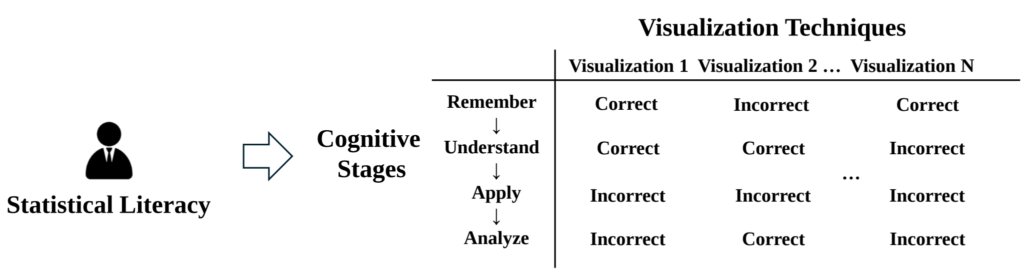

Figure 1: Analytical framework for mapping comprehension items across cognitive stages and visualization techniques

2.1 Statistical Literacy Measurement

In the context of the empirical study, participants were asked to complete a statistical literacy assessment. To evaluate their literacy levels, this study adopted Watson's (1997) three-tiered hierarchy of statistical skills, which allows classification into the following three tiers:

- Tier 1: Basic understanding of probabilistic and statistical terminology
- Tier 2: Understanding of probabilistic and statistical concepts embedded in broader social contexts
- Tier 3: Critical attitude toward conclusions lacking sound statistical foundations

Various models of statistical literacy have been proposed, including Gal's (2002) and Garfield's (2003) frameworks. This study adopts Watson's (1997) hierarchical model, as it was originally developed and empirically validated with secondary school students, making it especially suitable for structured assessments. The model's tiered structure reflected knowledge typically acquired through secondary education (equivalent to junior and senior high school in Japan), and therefore aligned well with the educational background of our target participants.

To operationalize statistical literacy in a way that reflected the Japanese secondary education curriculum (MEXT, 2017; 2018), we further categorized the required abilities into the following five types:

- Ability 1: Understanding data distributions
- Ability 2: Identifying data variability and outliers
- Ability 3: Collecting and sampling data appropriately
- Ability 4: Analyzing variability and relationships in data
- Ability 5: Engaging in probabilistic thinking and statistical inference

One item was developed for each of the five abilities at each of the three tiers, resulting in a total of 15 items (Table 1). Participants' responses (correct or incorrect) were used to measure their overall statistical literacy level.

Table 1: Structure of the statistical literacy assessment items

	Ability 1	Ability 2	Ability 3	Ability 4	Ability 5
Tier 1	Q1	Q2	Q3	Q4	Q5
Tier 2	Q6	Q7	Q8	Q9	Q10
Tier 3	Q11	Q12	Q13	Q14	Q15

2.2 Visualization Comprehension Measurement

This section describes how participants' comprehension is evaluated based on different ABSS visualization techniques. To this end, we used simulation results generated from Schelling's segregation model (Schelling, 1971) and presented them in five different visualization formats to assess comprehension levels. The Schelling model operates on a grid where agents (residents) follow a simple rule: they remain in place if the proportion of similar neighbors exceeds a given threshold, and move to an empty cell otherwise. As this process repeats, agents gradually cluster, resulting in emergent segregation patterns.

Because this model is intuitive and does not require specialized domain knowledge, it is well-suited for the purpose of evaluating comprehension of visualization techniques. In contrast, models tailored to specific fields, such as business or public policy, may produce results that depend heavily on participants' prior knowledge. To avoid such bias and to isolate the effect of visualization format alone, this study adopted Schelling's model as a neutral base.

2.2.1 Explanation of the Simulation Results

Before responding to questions, participants were provided with explanatory materials to facilitate understanding of the basic mechanisms and dynamics of Schelling's segregation model. These materials focused on the following points:

・Movement rules: Each agent checks the proportion of similar agents among the eight neighboring cells. If the proportion does not meet the threshold, the agent relocates to an empty space.
・Satisfaction threshold: Three different threshold settings (25%, 50%, and 75%) lead to distinct segregation patterns.
・Simulation process: For each threshold, 100 runs are conducted, and the final outcomes of segregation are recorded.

The explanatory material was made continuously available throughout the question-answering process to allow participants to consult it as needed for accurate interpretation.

2.2.2 Overview of Visualization Techniques

Simulation results from Schelling's model were represented using five different visualization techniques, and participants' comprehension of each was assessed. These visualization techniques were selected based on their frequent use in ABSS research, including common formats for representing agent distributions over time, as well as scenario analysis tools such as Landscape Analysis of Possibilities and Cladogram Analysis of Possibilities (Figure 2).

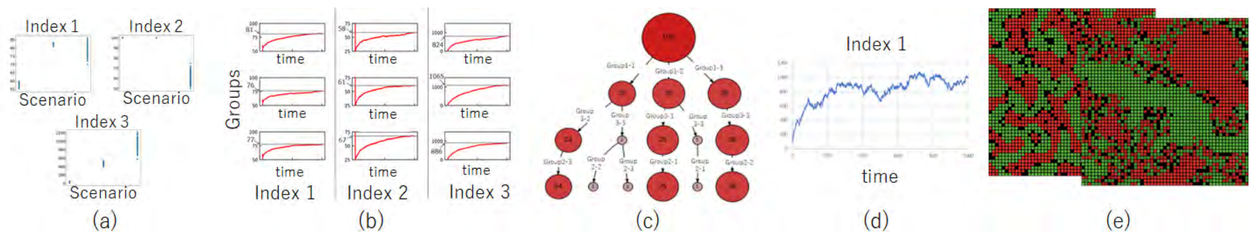

Figure 2: Visualization techniques used in the questionnaire

・Visualization 1: Landscape of Possibilities (Figure 2(a)) - This visualization plotted three outcome indicators after 1,000 simulation steps. The x-axis showed the satisfaction threshold, and the y-axis represented the indicator values. For each threshold, results from 100 simulation runs were plotted as blue dots.
・Visualization 2: Time-Series Clustering (Figure 2(b)) - This visualization clustered simulation runs into three groups based on similar trends over time. For each group, the average change in values over 1,000 steps was displayed.
・Visualization 3: Cladogram of Possibilities (Figure 2(c)) - This visualization showed the hierarchical relationships between the time-series clusters. The top-level node represented all 100 simulations, which branched into three groups: 26, 36, and 38 simulations, respectively.
・Visualization 4: Temporal Change (Figure 2(d)) - This visualization presented the time-series progression of a single simulation run, focusing on changes in the number of similar neighbors over time.
・Visualization 5: Agent Distribution (Figure 2(e)) - Agents were displayed as color-coded cells on a grid: red and green represented different types of agents, and black indicated empty cells. Although agents moved during the simulation, a static snapshot at a specific point in time was used for the questions.

Each visualization varies in its representational structure and level of visual complexity. Accordingly, the ease of comprehension may differ depending on users' statistical literacy and the cognitive stage they employ.

2.2.3 Design of Comprehension Questions

To evaluate participants' comprehension of visualizations, we developed items based on the revised version of Bloom's taxonomy (Anderson and Krathwohl, 2001). This revised taxonomy emphasizes action-oriented cognitive processes, Remember, Understand, Apply, Analyze, Evaluate, and Create. This verb-based structure provides for a clearer identification of the cognitive activities involved in learning and processing information, making it appropriate for assessing comprehension of visualized data. Although the revised taxonomy consists of six cognitive stages, this study focused on the first four: Remember, Understand, Apply, and Analyze. The remaining two stages, Evaluate and Create, involve advanced tasks such as evaluating model validity or generating new hypotheses, which require methodological expertise in ABSS and subjective judgment. These characteristics make them unsuitable for objective assessment in standardized questionnaires.

The four selected cognitive stages are defined as follows in the context of ABSS:

A. Remember: The ability to accurately read and extract numerical or factual information from visualizations.
B. Understand: The ability to interpret trends or relationships shown in the simulation outputs.
C. Apply: The ability to predict the effect of changes in simulation settings.
D. Analyze: The ability to infer causal relationships underlying the visualized outputs.

For each of the five visualization techniques, one item was created per cognitive stage, resulting in a total of 20 multiple-choice items (Table 2). Participants' responses (correct or incorrect) were used to assess the effectiveness of each visualization technique.

Table 2: Structure of visualization comprehension assessment items

	Visualization 1	Visualization 2	Visualization 3	Visualization 4	Visualization 5
Remember	A1	A2	A3	A4	A5
Understand	B1	B2	B3	B4	B5
Apply	C1	C2	C3	C4	C5
Analyze	D1	D2	D3	D4	D5

For example, the following were the items corresponding to each cognitive stage for the Landscape of Possibilities visualization:

・A1: How many simulation runs are plotted for each setting (25%, 50%, 75%)?
・B1: Which of the following best describes the segregation pattern trends for each threshold (25%, 50%, 75%) as shown in the Landscape of Possibilities figure?
・C1: If the satisfaction threshold is increased to 80%, how is the segregation pattern likely to change?
・D1: The Landscape of Possibilities figure shows that the variation across the 100 simulation runs differs by threshold. Which of the following best explains the reason for this variation?

3 DESCRIPTIVE ANALYSIS AND CORRELATION BETWEEN STATISTICAL LITERACY AND VISUALIZATION COMPREHENSION

The web-based survey was conducted in Japan from December 17 to December 23, 2024, yielding 392 valid responses. The participants ranged in age from 16 to 82 years, with a mean of 56.12 and a standard deviation of 13.11. The majority of participants were in their 40s to 70s. This section presents descriptive statistics on participants' statistical literacy levels and visualization comprehension scores (i.e., number of correct responses), followed by an analysis of the correlation between the two variables.

Table 3 shows the mean and standard deviation for each item used to measure statistical literacy. Scores for Tier 2 items (Q6–Q10), which assess the level of proficiency in interpreting statistics within broader social contexts, were relatively high, suggesting that participants possess a moderate level of proficiency to read and interpret statistical information in everyday or professional contexts. In contrast, scores for Tier 1 items (Q1–Q5), which focus on basic statistical concepts, were lower. This indicates that foundational knowledge, such as understanding averages and distributions, may not be well established among many participants. Rather than relying on formal or logical understanding, participants may be interpreting statistical information based on intuition or heuristic reasoning. This tendency is likely to be more pronounced among participants who completed their formal education many years ago, as they may have forgotten much of what they learned in school about basic statistics.

Table 3: Means and standard deviations for statistical literacy assessment items

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
mean	0.423	0.329	0.513	0.298	0.087	0.806	0.503	0.668	0.554	0.579	0.528	0.495	0.375	0.120	0.543
std	0.495	0.470	0.500	0.458	0.282	0.396	0.501	0.471	0.498	0.494	0.500	0.501	0.485	0.325	0.499

Next, Table 4 presents the mean and standard deviation for each item used to assess visualization comprehension. From the perspective of cognitive processes, the items related to Remember (A1–A5) showed relatively high accuracy, indicating that many participants were able to perform tasks involving the extraction of numerical information visually presented in the visualization. In contrast, scores for the Understand items (B1–B5) were generally low, suggesting that many participants struggled to grasp trends and relationships within the visualized data. Similarly, scores for the Apply items (C1–C5) were also low overall, indicating difficulty in predicting outcomes based on hypothetical changes. Finally, while the Analyze items (D1–D5) exhibited greater variability in correct response rates, the average scores remained at a moderate level.

Table 4: Means and standard deviations for visualization comprehension assessment items

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
mean	0.533	0.559	0.296	0.566	0.406	0.288	0.112	0.372	0.245	0.306	0.383	0.181	0.273	0.347	0.214	0.416	0.385	0.268	0.260	0.309
std	0.500	0.497	0.457	0.496	0.492	0.454	0.316	0.484	0.431	0.461	0.487	0.386	0.446	0.477	0.411	0.493	0.487	0.443	0.439	0.463

Figure 3 presents a scatterplot illustrating the relationship between the number of correct responses for statistical literacy and visualization comprehension. The size of each plot point visually represents the number of participants corresponding to each combination of scores. A Pearson correlation analysis yielded a coefficient of r = 0.623, indicating a moderate positive correlation between the two variables. This suggests that participants with higher levels of statistical literacy tended to interpret visualizations more accurately. However, the relationship was not strictly linear. As shown in Figure 3, there were a number of participants who, despite having high statistical literacy scores, exhibited low comprehension of visualizations. This indicates that differences in the components of statistical literacy, such as tier levels or specific abilities, may interact with the stages of cognitive processing and the types of visualization, resulting in varying effects on comprehension.

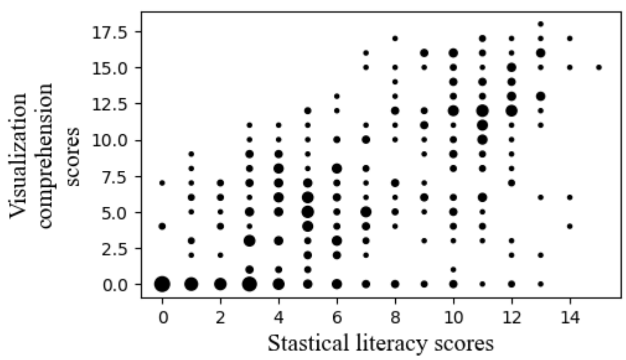

Figure 3: Relationship between statistical literacy and visualization comprehension scores

Table 5 presents the average number of correct responses for each cognitive stage, aggregated by participants' total scores on the statistical literacy assessment. Overall, as statistical literacy scores increase, visualization comprehension scores also tend to rise. However, the degree of this increase varies depending on the cognitive stage.

In the Remember stage, accuracy rates were the highest across all levels of statistical literacy, indicating that most participants were capable of accurately reading visual information regardless of their literacy level. In the Understand stage, participants with statistical literacy scores of 11 or below recorded the lowest comprehension scores, whereas those with scores of 12 or above performed second best after the Remember stage. This suggests that, while sufficient statistical knowledge enables participants to interpret visual structures and trends more effectively, many still experience difficulty at this stage. In the Apply stage, average scores remained low across all levels, suggesting that the hypothetical or predictive thinking required at this stage cannot be fully supported by basic statistical knowledge alone. For the Analyze stage, the scores were somewhat higher than those in the Apply stage, resulting in moderate overall performance. This implies that a certain number of participants were able to make plausible inferences about causal structures by using visual patterns as cues.

Table 5: Average visualization comprehension scores by statistical literacy level and cognitive stage

Correct Answers	A	B	C	D
0-2	0.692	0.231	0.692	0.404
3-5	1.672	0.776	0.897	1.216
6-8	2.347	1.000	1.213	1.627
9-11	3.286	2.041	2.071	2.337
12-	3.863	2.784	2.235	2.529

4 PREDICTIVE MODELING OF VISUALIZATION COMPREHENSION BASED ON STATISTICAL LITERACY

This section provides a more detailed analysis of the relationship between statistical literacy and visualization comprehension. Specifically, we examine which individual statistical literacy items (out of the 15 assessment items) most strongly influence the correctness of responses to visualization comprehension items, which are categorized by combinations of cognitive stages and visualization techniques. This analysis deepens the understanding of the overall trends presented in the previous section. The practical implications of these findings will be discussed in detail in Section 5.

To investigate this relationship, we conducted a machine learning-based analysis consisting of two steps: training classifiers to predict correctness on each visualization comprehension item based on participants' statistical literacy profiles, and examining the importance of each statistical literacy item in determining the prediction outcomes. In this approach, each visualization comprehension item was treated as a target variable (binary: correct or incorrect), while the 15 statistical literacy items were treated as explanatory variables. In the application of machine learning, high multicollinearity among explanatory variables can adversely affect both the predictive performance and interpretability of the model. As part of the evaluation, the Variance Inflation Factor (VIF) is computed to assess the degree of multicollinearity. The maximum value of VIF was found to be 1.71, indicating no multicollinearity concerns. Therefore, no variables were removed or merged. Among various classification models such as logistic regression, support vector machines (SVM), and decision trees, we selected LightGBM (Ke, et al. 2017), which provided the highest prediction accuracy in preliminary experiments. It should be noted that the choice of LightGBM was made solely to ensure predictive accuracy in this study, and we do not claim its general superiority. The primary goal of this analysis is to demonstrate the effectiveness of a multivariate framework for identifying how visualization comprehension varies depending on statistical literacy, cognitive stages, and visualization techniques. In future applications, other models may also be appropriate depending on the purpose and characteristics of the data.

For model training, we first optimized the LightGBM hyperparameters for each visualization comprehension item using the parameter ranges listed in Table 6. A five-fold cross-validation was performed, and the mean accuracy was used as the evaluation metric. Bayesian optimization was applied to identify the best-performing parameters.

After optimization, we trained classifiers using the selected parameters to predict participants' response correctness for each visualization comprehension item. The performance of each classifier was evaluated using accuracy as summarized in Table 7. Since the proportions of correct answers for some items were very low, we additionally evaluated the model using the F1-score as shown in Table 8. Some items yielded

F1-scores close to zero, indicating that statistical literacy had little explanatory power for those particular items.

Table 6: Parameter ranges for LightGBM optimization

Parameters	Min	Max
reg_alpha	0.0001	10
reg_lambda	0.0001	10
num_leaves	2	10
Colsample_bytree	0.1	1
Subsample	0.1	1
Subsample_freq	0	7
Min_child_samples	10	100

Table 7: Prediction accuracies for visualization comprehension items

	Visualization 1	Visualization 2	Visualization 3	Visualization 4	Visualization 5
Remember	0.735	0.745	0.709	0.709	0.709
Understand	0.771	0.888	0.732	0.781	0.745
Apply	0.724	0.819	0.730	0.658	0.786
Analyze	0.655	0.640	0.735	0.765	0.709

Table 8: F1-scores for visualization comprehension items

	Visualization 1	Visualization 2	Visualization 3	Visualization 4	Visualization 5
Remember	0.749	0.772	0.081	0.745	0.606
Understand	0.512	0.000	0.605	0.474	0.528
Apply	0.622	0.000	0.018	0.194	0.000
Analyze	0.546	0.402	0.021	0.390	0.392

According to Table 7, some items related to Visualizations 2 and 5 achieved relatively high accuracy rates. However, Table 8 reveals that certain items, specifically B2, C2, and C5, had F1-scores of zero. As shown in Table 4, these items had very low correct response rates in the original questionnaire data, and the classifiers predicted almost all responses as incorrect. This implies that these items were difficult regardless of participants' statistical literacy levels. In this way, the interpretation of accuracy and F1-scores suggests several implications.

From the perspective of visualization techniques, multiple cognitive stages were associated with low F1-scores for Visualizations 2 and 3. In particular, Visualization 3 resulted in low scores for Remember, Apply, and Analyze items. This suggests that Visualizations 2 and 3 hinder comprehension across several cognitive stages. One likely reason is the high complexity of these visualizations, which involve multiple interacting variables.

From the cognitive process perspective, Apply items (except for Visualization 1) consistently produced low scores. Although Visualizations 4 and 5 appear relatively simple, Apply items, requiring participants to predict changes resulting from simulation setting modifications, proved challenging for participants. Similarly, the F1-scores for Analyze items were also low, though not as severely as for Apply. These results suggest that reducing visual complexity alone is insufficient to support deeper levels of cognitive engagement such as Apply and Analyze.

Next, we analyzed which statistical literacy items serve as stronger predictors for each visualization comprehension item. Figure 4 presents the feature importance results from the LightGBM classifiers, with visualization comprehension items on the x-axis and statistical literacy items on the y-axis. In each column, greater values are represented by lighter shades. The heatmap shows that certain applied statistical literacy items, such as Q6, Q9, and Q14, were among the most important predictors across multiple comprehension items. This finding suggests that applied aspects of statistical literacy play a more significant role in visualization comprehension than basic knowledge alone.

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5
Q1	12	9	48	33	28	15	25	55	90	21	52	6	492	65	28	2197	552	4	84	53
Q2	102	62	101	51	47	8	4	37	86	383	63	62	73	52	0	7	166	25	48	34
Q3	26	57	53	24	24	123	3	68	53	25	38	10	23	5	35	5498	32	41	31	19
Q4	39	119	3	39	53	39	47	56	111	32	39	63	18	24	16	77	0	358	60	391
Q5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q6	21	306	4	10	424	23	10	542	13	39	195	6	7	62	30	5477	56	14	101	62
Q7	70	63	47	32	45	4	28	61	3	57	45	35	21	40	14	36	39	11	7	12
Q8	15	34	41	57	7	106	0	63	1	7	3	13	1	73	105	73	11	44	18	161
Q9	8	10	70	324	40	483	309	16	57	11	11	1	67	6	90	4	5	10	58	19
Q10	31	45	36	20	20	5	22	20	4	13	463	398	14	40	23	30	18	12	37	47
Q11	12	9	40	23	62	10	32	23	6	44	14	78	11	8	15	114	84	44	1	70
Q12	520	110	63	64	11	22	1	25	69	17	45	18	67	73	37	85	6	48	37	82
Q13	70	76	12	45	22	50	76	43	37	36	69	61	10	68	31	17	19	31	59	6
Q14	121	0	448	204	155	88	0	470	86	98	0	81	416	0	355	0	737	12	475	82
Q15	30	70	72	41	49	83	78	33	398	11	29	4	58	31	45	48	38	48	10	37

Figure 4: Feature importance of statistical literacy items for predicting visualization comprehension

On the other hand, for comprehension items associated with Apply and Analyze stages, items such as Q1-Q4, which assess basic statistical concepts, emerged as important predictors. This implies that fundamental statistical knowledge is essential for reasoning about the underlying information presented in visualizations. In particular, Visualization 1 required a solid understanding of basic statistics, as it closely resembles a scatterplot. Without the ability to interpret such visuals statistically, participants would struggle to perform accurate analysis.

In summary, this analysis reveals that statistical literacy is not a monolithic skill but rather a collection of abilities, each influencing visualization comprehension in different ways. Furthermore, the unequal distribution of feature importance across the 15 statistical literacy items suggests the potential to predict comprehension using a smaller subset of items, which could improve the efficiency of future assessments.

5 PRACTICAL IMPLICATIONS FOR ABSS VISUALIZATION

Based on the analyses presented in Sections 3 and 4, this section outlines practical implications for selecting visualization techniques in ABSS. The findings suggest that different cognitive stages, Remember, Understand, Apply, and Analyze, require distinct considerations when choosing appropriate visualizations depending on ABSS users' statistical literacy levels. The following summarizes key insights for each stage.

First, at the Remember stage, the relatively high F1-scores indicate that ABSS users’ ability to comprehend visualizations can be predicted by their statistical literacy. As shown in Table 4, the correct response rates for Remember items were high overall, suggesting that ABSS users with a certain level of statistical literacy can perform basic data extraction without additional visual aids. However, Visualization 3, due to its structural complexity, yielded low F1-scores and low accuracy, indicating that it is difficult to comprehend regardless of literacy level. Such complex visualizations should ideally be avoided.

At the Understand stage, although the F1-scores were not as high as those at the Remember stage, they were still relatively high. Some items also showed clear differences in comprehension depending on statistical literacy levels. This suggests that this stage offers useful cues for selecting appropriate visualization techniques based on ABSS users’ literacy. However, as Table 5 shows, participants with lower literacy scores struggled significantly at this stage. This implies that difficulties may arise regardless of visualization technique, requiring careful selection. If ABSS users fail to understand information at this stage, subsequent stages (Apply and Analyze) become less meaningful. Furthermore, Visualization 2 had an F1-score of 0, and the proportion of correct responses among participants was also low, indicating that it is difficult to comprehend regardless of statistical literacy. Since this visualization aggregates multiple time series into clustered averages, the original meaning becomes difficult to grasp. At the Understand stage, visualizations must emphasize simple structure and clearly perceptible temporal or relational changes.

At the Apply stage, all visualizations except for Visualization 1 had extremely low F1-scores, indicating that comprehension at this stage cannot be predicted based on statistical literacy. Even participants with 12 or more correct responses on the literacy assessment had the lowest accuracy at this stage (Table 5). This suggests that even ABSS users with strong statistical backgrounds may find it difficult to reason through the Apply stage in ABSS contexts. Therefore, those who present ABSS outputs must recognize that many ABSS users may face challenges here, regardless of how carefully visualizations are selected. This challenge may stem from the need to understand the causal dynamics resulting from changes in simulation settings. Simplifying the visuals alone is insufficient; it is likely necessary to incorporate narrative elements that explain cause-and-effect changes more explicitly.

Finally, at the Analyze stage, although the challenge is not as severe as in Apply, it remains difficult to predict effective visualization techniques based solely on statistical literacy. However, Table 5 shows that over half of the participants with 12 or more correct answers on the literacy assessment were able to answer Analyze items correctly. This suggests that when ABSS users possess a high level of statistical literacy, effective visualization selection may still be possible. Furthermore, items Q1 to Q4, covering fundamental statistical concepts such as averages and distributions, were found to be important for selecting appropriate visualizations. Therefore, assessing ABSS users' comprehension of these basic concepts may help determine whether certain visualization techniques are suitable.

In summary, this study suggests that selecting visualization techniques for ABSS outputs can benefit from an awareness of ABSS users' statistical literacy levels and the cognitive stages involved. However, existing visualization techniques may be insufficient to support comprehension at the Apply stage. Thus, supporting ABSS user comprehension in such cases likely requires additional methods beyond visualization alone, such as narrative explanations or other forms of cognitive support.

6 CONCLUSION

This study proposed a structured analytical framework for evaluating the comprehension of ABSS visualizations, grounded in ABSS users' statistical literacy and cognitive stages. It then applied this framework in an empirical study to explore how different visualization techniques support comprehension at various cognitive stages. In particular, the study focused on four stages of cognitive processing, Remember, Understand, Apply, and Analyze, and examined the effectiveness of different visualization techniques at each stage and their relationship with statistical literacy.

The results revealed that certain cognitive stages allow for the prediction of effective visualization techniques based on ABSS users' statistical literacy. This suggests that it is possible to design visualizations that better support comprehension when ABSS user literacy levels are known. However, in the Apply stage, even ABSS users with high statistical literacy had difficulty interpreting the visualizations, indicating the need for supplementary methods beyond traditional visual representation.

6.1 Contributions

This study is the first in ABSS to incorporate insights from educational psychology to structure visualization comprehension into four cognitive stages, Remember, Understand, Apply, and Analyze, and to systematically assess comprehension at each stage. In doing so, it provides a novel theoretical and empirical foundation for evaluating the "intelligibility" of visualizations, traditionally discussed in subjective terms.

From a practical perspective, the study offers actionable insights to support the selection of visualization techniques when explaining ABSS outputs to ABSS users, based on their levels of statistical literacy. These findings can serve as a foundational framework for designing effective visualizations tailored to ABSS user capabilities. In fact, the application of ABSS has been expanding across various domains, including business fields such as finance, consumer markets, supply chains, and energy (Macal and North, 2010). Additionally, a wide range of tools is now available to support ABSS modeling and analysis (Abar et al., 2017).

In such contexts of social implementation, it becomes increasingly important to understand how ABSS users interpret simulation outputs derived from models and tools, and how these interpretations inform decision-making. Thus, beyond technical visual representation itself, this study highlights the necessity of considering how information can be presented in ways that support human comprehension and decision support. In Japan, Societal Prototyping Design (SPD) initiatives based on ABSS have already begun in several municipalities under national-level projects (JST, 2024). In these efforts, it is a growing challenge for local government staff and residents to correctly interpret the simulation results.

In this broader context, the present study provides practical knowledge for designing more effective communication of ABSS outputs to diverse ABSS users. It also contributes to the field of science communication, particularly research on knowledge transmission through data visualization (Zallio, 2021; Yang et al., 2019), by offering insights into how expert-generated knowledge can be effectively conveyed to non-expert users.

6.2 Future Directions

This study suggests three major directions for further development. First, empirical validation in specific domains is essential. While this study adopted Schelling's segregation model as a domain-independent simulation to minimize the influence of prior knowledge, real-world contexts such as policy decision-making and business environments often involve domain-specific expertise that can significantly affect the ABSS users' comprehension. Future research should use ABSS outputs grounded in real-world problems to evaluate how ABSS users with different roles or interests interpret visualizations, thereby identifying specific barriers to comprehension.

Second, new visualization methods need to be designed based on user characteristics. This study evaluated the effectiveness of existing visualization techniques, but going forward, it will be essential to develop adaptive visualization approaches tailored to ABSS users' levels of statistical literacy. In particular, for higher-order cognitive processes such as Apply and Analyze, interactive features or narrative-based visualizations that explain changes in simulation settings as coherent stories (Hullman and Diakopoulos, 2011; Botsis et al., 2020) may be especially effective in enhancing comprehension among ABSS users.

Third, it is important to integrate these insights into a practical guideline for selecting appropriate visualization techniques. As the social implementation of ABSS continues to expand, there is a growing need for comprehensive design principles that address the question of which types of visualizations should be presented to ABSS users with varying levels of statistical literacy and domain knowledge. The present study represents a first step toward that goal. Future work should aim to establish a support framework for visualization design that is both empirically grounded and theoretically robust, enabling practical application in real-world decision-making contexts.

ACKNOWLEDGMENTS

This work was supported by the JST-Mirai Program, Grant Number JPMJMI23B1.

REFERENCES

Abar, S., G. K. Theodoropoulos, P. Lemarinier, and G. M. P. O'Hare. 2017. "Agent-Based Modelling and Simulation Tools: A Review of the State-of-Art Software". *Computer Science Review* 24:13–33.

Anderson, L. W., and D. R. Krathwohl. 2001. *Taxonomy for Learning, Teaching, and Assessing*. New York: Longman.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. New York: David McKay Company.

Botsis, T., J. E. Fairman, M. B. Moran, and V. Anagnostou. 2020. "Visual Storytelling Enhances Knowledge Dissemination in Biomedical Science." *Journal of Biomedical Informatics* 106:103458.

Burns, A., C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. 2020. "How to Evaluate Data Visualizations across Different Levels of Understanding". In *Proceedings of the 2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, October 25th-30th, Virtual Event, 19–28.

Gal, I. 2002. "Adult's Statistical Literacy: Meanings, Components, Responsibilities". *International Statistical Review* 70(1):1–25.

Garfield, J. 2003. "Assessing Statistical Reasoning". *Statistics Education Research Journal* 2(1):22–38.

Gilbert, N. 2008. *Agent-Based Models*. Thousand Oaks, CA: Sage Publications.

Goto, Y. 2020. "Hierarchical Classification and Visualization Method of Social Simulation Logs Reflecting Multiple Analytical Interests". *Transactions of the Society of Instrument and Control Engineers* 56(10):463–474.

Goto, Y., and S. Takahashi. 2011. "Landscape Analysis of Possible Outcomes". In *Agent-Based Approaches in Economic and Social Complex Systems VI*, edited by S. Chen, T. Terano, and R. Yamamoto, 165–178. Tokyo, Japan: Springer.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". *Advances in Neural Information Processing Systems* 30:3149–3157.

Hullman, J., and N. Diakopoulos. 2011. "Visualization Rhetoric: Framing Effects in Narrative Visualization". *IEEE Transactions on Visualization and Computer Graphics* 17(12):2231–2240.

Japan Science and Technology Agency (JST). 2024. Development of Digital Social Experimentation Platform Technology Enabling Human-Centered Social Co-Creation Design. JST Future Society Creation Program. https://www.jst.go.jp/mirai/en/program/super-smart/JPMJMI23B1.html, accessed 11th April 2025.

Lee, J.-S., T. Filatova, A. Ligmann-Zielinska, B. Hassani-Mahmooei, F. Stonedahl, I. Lorscheid, A. Voinov, G. Polhill, Z. Sun, and D. C. Parker. 2015. "The Complexities of Agent-Based Modeling Output Analysis". *Journal of Artificial Societies and Social Simulation* 18(4):4.

Macal, C. M., and M. J. North. 2010. "Tutorial on Agent-Based Modelling and Simulation". *Journal of Simulation* 4(3):151–162.

Maltese, A. V., J. A. Harsh, and D. Svetina. 2015. "Data Visualization Literacy: Investigating Data Interpretation Along the Novice–Expert Continuum". *Journal of College Science Teaching* 45(1):84–90.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). 2017. Junior High School Course of Study: Mathemati cs. https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387018_004.pdf, a ccessed 11th April 2025.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). 2018. High School Course of Study: Mathematics. https://www.mext.go.jp/content/20230217-mxt_kyoiku02-100002620_05.pdf, accessed 11th April 2025.

North, M. J., and C. M. Macal. 2007. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford: Oxford University Press.

Padilla, L. M., G. Hansen, I. T. Ruginski, H. S. Kramer, W. B. Thompson, and S. H. Creem-Regehr. 2015. "The Influence of Different Graphical Displays on Nonexpert Decision Making Under Uncertainty". *Journal of Experimental Psychology: Applied* 21(1):37–46.

Schelling, T. C. 1971. "Dynamic Models of Segregation". *Journal of Mathematical Sociology* 1(2):143–186.

Schonlau, M., and E. Peters. 2008. "Graph Comprehension: An Experiment in Displaying Data as Bar Charts, Pie Charts and Tables with and without the Gratuitous 3rd Dimension". WR-618, RAND Corporation.

Tanaka, Y., M. Kunigami, and T. Terano. 2018. "What Can Be Learned from the Systematic Analysis of the Log Cluster of Agent Simulation". *Simulation & Gaming* 27(1):31–41.

Watson, J. 1997. "Assessing Statistical Thinking Using the Media". In *The Assessment Challenge in Statistics Education,* edited by I. Gal and J. B. Garfield, 107–121. Amsterdam: IOS Press.

Yamada, H., S. Yamane, K. Ohori, T. Kato, and S. Takahashi. 2021. "A Method for Micro-Dynamics Analysis Based on Causal Structure of Agent-Based Simulation". In *2020 Winter Simulation Conference (WSC),* 313–324. https://doi.org/10.1109/WS C48552.2020.9384118.

Yang, S., K. Lo, L. L. Wang , and J. West. 2019. "Delineating Knowledge Domains in the Scientific Literature Using Visual Information". *arXiv preprint arXiv:1908.07465*.

Zallio, M. 2021. "Democratizing Information Visualization: A Study to Map the Value of Graphic Design". In *Design, User Experience, and Usability: UX Research and Design*. DUXU 2021. Lecture Notes in Computer Science, vol. 12779, edited by A. Marcus, 495–508.

AUTHOR BIOGRAPHIES

KOTARO OHORI is a Professor in the Faculty of Business Administration at Toyo University. He received his Ph.D. from Waseda University. His current research interests include social system design based on artificial intelligence and agent-based social simulation. His email address is ohori@toyo.jp.

KYOKO KAGEURA is an undergraduate student in the Faculty of Business Administration at Toyo University. Her email address is s13102203293@toyo.jp.

SHOHEI YAMANE is a Researcher in the Fujitsu Research. He earned his Ph.D. in Information Science from Kyoto University. His current research interests include artificial intelligence and agent-based social simulation. His email address is yamane.shohei@fujitsu.com.