A Study on the Designation Institution for Supercomputer Specialized Centers in Republic of Korea

Hyungwook Shim, Yonghwan Jung, Jaegyoon Hahn*
National Supercomputer Department, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea

Abstract—In Korea, specialized centers are designated for 10 strategic fields for the purpose of jointly utilizing supercomputer resources at the national level. Based on the “National Supercomputing Innovation Strategy,” it plans to select 10 centers in three stages by 2030, and has now completed the designation of the first-stage specialized centers in 2022. With the second designation in 2024 ahead, it is urgent to review and improve the existing designation institution for fairer and more effective selection of specialized centers. Therefore, this paper analyzed the influence of evaluation items and the influence of evaluation items on evaluation results by using logistic regression analysis and network centrality analysis to prepare improvement plans for the existing evaluation model. As a result of the analysis, improvement measures were derived, such as subdividing evaluation items with low impact, expanding the items, and lowering the allotment of evaluation items with low impact.

Keywords—Supercomputer; specialized center; evaluation system; logistic regression model; network centrality analysis

I. INTRODUCTION

Korea’s supercomputer governance consists of a national center, a specialized center, and a unit center. The national center secures and operates supercomputing resources, supports policy establishment, and manages joint utilization. The specialized center performs supercomputing resource establishment and operation, basic application research and dissemination of results, etc. The unit center is a resource independently operated by individual private research institutes and companies [1]. Specialized centers maintain their qualifications for five years after designation. Currently, designation of specialized centers in 7 fields has been completed, and by 2028, it will be expanded to 10 fields[2]. Recently, the Ministry of Science and ICT announced that it would establish the “3rd supercomputer development basic plan” (referred to as ‘3rd Basic Plan’), the top plan for supercomputers, and establish a user support system centered on specialized centers. Therefore, at the beginning of the “3rd Basic Plan”, the government should make efforts to improve the special center designation institution to ensure fairness, effectiveness, and sustainability. As a measure to improve the government's evaluation system in the field of science and technology, a statistical method using the influence of each evaluation item on the evaluation result is widely used [3]. Therefore, this paper also presents a plan to improve the specialized center designation institution by using the evaluation results for the designation of the existing specialized center.

This paper consists of six sections. Sections I and II presented an academic value through a qualitative analysis of the background, meaning, and source research of this paper. Section III introduces the function, role, and protection system of the supercomputer specialized center, and Section IV explains the methodology of this paper. In Section V, a case study for improving the evaluation model is conducted and the results are gradually presented. Finally, in Section VI, the results were summarized and the viewpoint was straightened out, and the final point and pursuit plan of this paper were presented.

II. LITERATURE REVIEW

Major prior studies are as follows. Hirao (2010) introduced projects for the introduction of peta-class next-generation supercomputing systems [4], and Hsu (2015) analyzed foreign trends for exascale supercomputing development and introduced major projects invested in the United States [5]. Mitsusha (2021) introduced Fugaku's flagship project related to Japan's Fugaku supercomputer and presented design details such as Fugaku's scale and performance [6]. Savin (2019) introduced the supercomputing center community system in Russia and mentioned the advantages in terms of energy efficiency and the provision, monitoring, and management of resources through a shared utilization network. In addition, improvement plans were presented through analysis of the current status of the Joint Supercomputer Center [7]. Prior domestic studies are as follows. Huh (2021) conducted research on ways to improve the legal system to vitalize the supercomputing ecosystem in Korea. Regarding the supercomputer-related law, the ‘Supercomputer Act’, problems such as the role of related institutions, project costs, mutual cooperation system, and consistency with higher-level plans were identified, and improvement measures were proposed with a focus on policy consistency and effectiveness enhancement [8]. Shim (2022) conducted a study to improve the evaluation index for selecting a research institute for the national R&D project of the Ministry of Land, Infrastructure and Transport. Using the evaluation score of each evaluator, the evaluation index was determined through an artificial neural network, and a method for improving the score distribution for each evaluation index was derived using logistic regression analysis [9]. Shin (2013) conducted a study to prepare improvement plans for local government
performance indicators, classified performance indicators by characteristics, and applied development procedures that considered visions and promotion strategies for each local government and key outcomes by function. In addition, improvement measures were derived considering desirable performance indicator attributes such as relevance, clarity, timeliness, reliability, and comparability [10]. Lee (2018) set the field and elements for improving the educational environment and set the direction of the indicators through a delphi survey by experts for the purpose of developing educational environment improvement indicators. As a result of the analysis, the indicators for improving the educational environment, such as the adequacy of the total floor area of the classroom, the adequacy of general school teaching, whether or not to secure seismic performance, the deterioration of firefighting facilities, and energy consumption, were finally determined [11]. Ji (1999) conducted research for the rational development of informatization indicators that measure the national informatization level. The author first set informatization facility indicators, informatization use indicators, and informatization support indicators, subdivided them into 6 groups, and proposed an informatization measurement indicator system consisting of a total of 28 indicators in consideration of informatization level, reality, and applicability aspects [12]. Kim (2022) conducted a study on how to improve the global cyber security index, which is used to diagnose the level of national cyber security development and strengthen cyber security capabilities. The author established basic principles for the improvement and utilization of the Global Cyber Security Index and suggested development plans through survey-based SWOT analysis [13].

The academic value of this paper is as follows. It is novel because no research has been conducted on the improvement of evaluation system related to the existing domestic supercomputer. Although Huh (2021) conducted a study on institutional improvement measures related to supercomputers, this paper only examines the appropriateness of the evaluation system. Also, research related to the development and improvement of existing evaluation system draws conclusions using a qualitative method of asking and organizing the opinions of experts. However, in this paper, the influence of each index of the existing evaluation model was analyzed using a statistical method using the actual evaluation result data of the evaluation committee, and a comparative analysis was performed with the improved evaluation model. Lastly, a survey was conducted on the appropriateness of the evaluation index targeting the researchers of the specialized center support institution that was evaluated, and the fairness of the improvement model was added by reflecting the opinions of all parties participating in the evaluation.

III. SUPERCOMPUTER SPECIALIZED CENTER

A. Definition of Supercomputer Specialist Center

Supercomputer specialized center defined as an institution that possesses expertise for professional use of supercomputers, provides specialized services based on resources, manpower, and technology specialized in the field, conducts research and development, and promotes the use of supercomputers. The functions and roles of the specialized centers in Table I include the establishment and operation of supercomputing resources by field, service provision, base application research and dissemination of research results, large-capacity data management and operation support, and human resource training.

B. Evaluation Institution for Designation of Supercomputer Specialized Center

The designation of specialized centers is in accordance with the "Operational Guidelines for Designating Supercomputers by Field" (referred to as ‘Operational guidelines’). The Operating guidelines are based on Article 9-2 of the “Act on the Promotion and Utilization of Supercomputing” (referred to as the ‘Supercomputer Act’), which includes the functions and roles of specialized centers, designation procedures and methods, establishment of operation plans, evaluation of operational performance, composition and operation of evaluation teams, etc. The designation procedure is shown in Fig. 1. First of all, the institutes, universities target organizations that can submit applications are central administrative agencies, national and public research, private companies and organizations with expertise in each field, and meet the requirements such as 4 or more supercomputer experts and supercomputer possession (1.5 million dollars or more). After submitting the application, one institution in each field can be designated as a specialized center through the first written examination, the holding of a briefing session, and the second face-to-face examination. Finally, among institutions with a score of 70 or more, the institution with the highest score is selected.

In the evaluation system, the subject of evaluation is the evaluation team, and it is made up of 3 or more and 10 private experts, including the head. The evaluation team conducts both the first and second evaluation, and the main evaluation items are shown in Table II. The items and indicators were derived through the FGI of experts in the related field, and the evaluation items include the ‘Related Performance’, ‘Validity of operation purpose and plan’, ‘Suitability of center manpower’, etc.

<table>
<thead>
<tr>
<th>TABLE I. FUNCTIONS AND ROLES OF THE SPECIALIZED CENTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function and Role</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>• Providing supercomputing services in specialized fields</td>
</tr>
<tr>
<td>• Research and development of supercomputing technology and application specialized in specialized fields</td>
</tr>
<tr>
<td>• Expansion of resource sharing through participation in the joint utilization system</td>
</tr>
<tr>
<td>Service</td>
</tr>
<tr>
<td>• Establishment, operation and service of specialized resources for each sector centered on 10 strategic fields</td>
</tr>
<tr>
<td>• Utilization support and consulting through field experts</td>
</tr>
<tr>
<td>• Community development and education by field</td>
</tr>
</tbody>
</table>

Fig. 1. Specialized center designation procedure.
In the specialized center designation institution, the evaluation items and indicators for designation are important factors in selecting a specialized center representing a specific field for the next five years. Regarding the national R&D project evaluation system, Lee (2010) also emphasizes the effects of the need for improvement of evaluation items and indicators, and the preparation of improvement plans using evaluation results [14]. Using the actual evaluation result data, the appropriateness of new discovery of evaluation items, removal of existing items, and adjustment of points allocation was proved. Therefore, in this paper, it is necessary to review the appropriateness of evaluation items based on the evaluation results of the seven specialized centers.

IV. THEORETICAL BACKGROUND AND METHODOLOGY

A. Research Procedure

In many papers such as Ahn(2022), regression analysis, AHP, and machine learning analysis are mainly used to research measures to improve the evaluation system [15]. The analysis results are effective in discovering individual improvement factors for evaluation items, and can provide intuitive results to researchers. However, evaluation items are grouped into various sub-indicators, and due to the nature of R&D, the range of evaluation items such as research method, content, and research timing is radially intertwined. Therefore, it is more effective to discover improvement factors by considering the correlation between evaluation items, and in this paper, an improvement plan was derived considering the result of network centrality analysis to reflect the correlation between evaluation items. The research procedure of this paper is shown in Fig. 2. First, using logistic regression analysis for the existing evaluation model, the influence on the selection result for each evaluation item and the appropriateness of the points assigned are reviewed. Second, through network centrality analysis, the evaluation model improvement plan is derived by analyzing the structure and centrality that affect evaluation items. Finally, the results are summarized and implications are drawn.

B. Logistic Regression Model Analysis

Binary logistic regression analysis is used to estimate the relationship between a binary dependent variable (categorical) and multiple independent variables to explain the influence of the independent variable on the dependent variable or to predict the value of the dependent variable for the value of the independent variable. It is an analysis technique that regression models can generally be expressed as in Equation (1), and y* has values of 0 and 1.

\[
\hat{y} = b_0 + b_1X_1 + b_2X_2 \cdots b_kX_k + \epsilon \tag{1}
\]

As shown in Fig. 3, the regression model can be divided into total variance SST (Total Sum of Squares), which means the difference between the actual value and the mean, and error variance SSE (Error Sum of Squares), which means the difference between the actual value and the estimated value, and variance by the regression equation SSR (Sum of Squares due to Regression). Through these three fluctuation values, the coefficient of determination, which means the contribution to explain the diversity of the dependent variable, is obtained, and the regression coefficient b can be estimated as shown in Equation (2) through the least squares method in which the variation in error(SSE) is minimized.

\[
b_k = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}
\]

In this paper, the selection result was substituted as a dependent variable, and ‘0’ was set as not selected and ‘1’ as selected. Independent variables represent 4 evaluation items excluding added points, and were analyzed by standardizing them to values between 0 and 1.

![Fig. 2. Specialized center designation procedure.](image2)

![Fig. 3. Regression model conceptual diagram.](image3)

<table>
<thead>
<tr>
<th>Evaluation item</th>
<th>Evaluation indicator</th>
<th>points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related Performance</td>
<td>Performance of R&amp;D manpower training</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Performance of service, data management and research support</td>
<td>10</td>
</tr>
<tr>
<td>Validity of operation purpose and plan</td>
<td>Justification and Necessity of Designation</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Challenge and specificity of vision and operational goals</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Suitability of goals, project contents, research methods, etc.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Excellence of Expected Performance and Utilization Plan</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Center fostering and operation support plan</td>
<td>10</td>
</tr>
<tr>
<td>Suitability of center manpower</td>
<td>Appropriateness of personnel planning and organizational composition</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Researcher’s expertise</td>
<td>10</td>
</tr>
<tr>
<td>Facility and Equipment Securing Plan</td>
<td>Facility, Equipment Status and Expansion Plan</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Facility Status and Expansion Plan</td>
<td>5</td>
</tr>
<tr>
<td>Added points</td>
<td>Ratio of joint utilization resources</td>
<td>10</td>
</tr>
</tbody>
</table>
C. Network Centrality Analysis

Network analysis is an analysis method that analyzes the characteristics of relationships that exist between objects (nodes). It can visualize a microscopic network of relationships by deriving characteristics such as connection strength and connection structure between research subjects using nodes and links. Network centrality analysis is an analysis method that utilizes centrality indicators using the number of connections between nodes, distances, and travel routes among various measurement indicators [16]. Connectivity centrality, proximity centrality, betweenness centrality, and eigenvector centrality are typically used as centrality indicators [17]. The eigenvector centrality index used in this paper is a method of analyzing centrality by weighting the centrality of the other node. Characteristically, as shown in Equation (3), the maximum eigenvalue \( \lambda \) of the matrix between nodes is used.

\[
C_i = \frac{1}{\lambda} \sum_{j} a_{ij} C_j
\]

(3)

\( i,j \): Node

C: Eigenvector centrality

\( \lambda \): eigenvalue

\( \alpha \): Connection relationship (connected 1, unconnected 0)

Nodes mean evaluation items during analysis, and using the matrix constructed using the survey results, eigenvalues are calculated and eigenvector centrality for evaluation items and indicators is estimated [18].

V. Case Study

A. Data

Data for case studies can be classified into two types. First, the data for the logistic regression analysis use the evaluation result data of the specialized center evaluation team. The evaluation result data uses written evaluation data for 7 fields, and uses the evaluation scores written by the evaluation team for a total of 15 institutions in 7 fields. The dataset is shown in Table III. A, B, C, and D were selected as independent variables and entered as 0~1 scale. The dependent variable was selected as a nominal variable (dummy variable) with two values of 0 and 1 depending on whether or not it was selected.

Next, as data for network centrality analysis, data from an online survey targeting 60 people, including executives and employees belonging to seven specialized centers, are used. The survey items consist of a total of 17 items, including a total of 16 items in the Likert scale method and one item in the matrix method.

B. Logistic Regression Analysis Results

The analysis results of the regression model including all evaluation items are shown in Tables IV and V. Table IV shows the Nagelkerke R2 index, which indicates the explanatory power of the entire model, and the Hosmer-Lemeshow index, which is a goodness-of-fit test of the model. In the Hosmer-Lemeshow test, the chi-square value indicates the degree of agreement between the actual value of the dependent variable and the predicted value by the model and the smaller the chi-square value, the higher the fitness of the model. The extent to which independent variables explained the dependent variable was about 24%. Since the value of the significance probability \( p \) in the goodness of fit of the model was larger than 0.05, the null hypothesis was not rejected, so the goodness of fit can be considered acceptable.

Table V is the estimation result of the regression coefficient of the model. First, if the sign of the regression coefficient \( \beta \) is positive (+), the greater the value of the corresponding independent variable, the greater the possibility of being classified as a selected group representing the dependent variable ‘1’, and negative (-) means the opposite case. As a result of the analysis, evaluation items A and D are positive (+), and the higher the score of evaluation items A and D, the higher the possibility of being selected. B and C are negative (-), and the higher the score, the higher the possibility of not being selected. The significance probability is less than 0.05 for both evaluation items A and D in the 95% confidence interval, which can be considered significant, and B and C are 0.227 and 0.901, respectively, which are greater than 0.05, so it can be considered insignificant. Wald is a statistic that verifies whether the coefficient value for each covariate is zero. \( \exp(\beta) \) represents the odds ratio and means the influence on the evaluation result when an evaluation item increases by one unit. It can be interpreted that when the score of evaluation item A increases by one unit, the probability of being selected increases about 20 times.
As a result of the statistical analysis of Table VI, it can be confirmed that the evaluation items B and C are not appropriate in the direction and significance probability affecting the evaluation result. In the case of the significance probability, it cannot be a factor that absolutely determines the validity of the independent variable, but in the case of the direction, it can be a factor that can determine the validity of the independent variable in consideration of the evaluation criteria. Therefore, after excluding the two evaluation items with a negative (-) sign, the regression coefficient was re-estimated, and the re-estimation results are shown in Table VII. The regression coefficients of the A and B evaluation items showed positive (+) values, and it was confirmed that they were somewhat reduced compared to the previous ones. The significance probability was statistically significant at the 90% confidence interval, and it was confirmed that the influence of the evaluation result was somewhat lowered according to the odds ratio result. The influence of D on the selection result was about 1.6 times greater than that of A. Through this, it is necessary to improve evaluation items B and C as a method for modifying the model. In order to prepare improvement measures, the network centrality analysis results are additionally conducted and the two analysis results are comprehensively considered.

### TABLE VI. RESULT OF THE REGRESSION COEFFICIENT ESTIMATION

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E</th>
<th>Wald</th>
<th>Significance Probability</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.638</td>
<td>.918</td>
<td>3.188</td>
<td>.074</td>
<td>5.147</td>
</tr>
<tr>
<td>D</td>
<td>2.102</td>
<td>1.080</td>
<td>3.788</td>
<td>.052</td>
<td>8.184</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.297</td>
<td>1.780</td>
<td>12.515</td>
<td>.000</td>
<td>.002</td>
</tr>
</tbody>
</table>

### C. Network Centrality Analysis Result

The result of visualizing the network for evaluation items is shown in Fig. 4. Node A represents relevant performance, node B represents the validity of operation purpose and plan, node C represents the suitability of center manpower, node D represents the facility and equipment securing plan, and node E represents the evaluation items for the add points. The size of a node increases as the frequency of the node increases, and the frequency is determined by the number of choices made by the respondent. The link is expressed as a straight line connecting the nodes, and the higher the co-occurrence frequency, the bolder it is.

![Fig. 4. Regression model conceptual diagram.](image)

As a result of the analysis, the frequency of nodes was highest in evaluation item B, which means the validity of the operation purpose and plan, followed by D, E, A, and C in order. For the links, the B-D link that connects the validity of the operation purpose and plan and the add points evaluation items appeared in the thickest form, followed by B-C and D-E in that order. In other words, it can be analyzed that the strength of the relationship between B-D, B-C, and D-E is relatively strong and the strength of the relationship between the evaluation items C-A and A-E is relatively weak around evaluation item B in the network between evaluation items evaluated by the response group. The strength of the relationship can be interpreted in various ways, but from the perspective of the evaluation system, it can be interpreted in two cases. First, the allocation of evaluation items with high relationship strength should be relatively higher than those with low relationship strength. This is because the higher the relationship strength, the greater the effect on the overall evaluation scores. Second, it is necessary to distinguish between evaluation item groups with high relationship strength and evaluation item groups with low relationship strength. If the impact on the evaluation results is significantly large due to the difference in relationship strength, it may be appropriate to classify or exclude groups of evaluation items. Therefore, in this paper, when considering the improvement of the overall evaluation system, the relationship strength was used as a basis for adjusting the evaluation item. In order to analyze the quantitative influence of each evaluation item in the network, the centrality value is used, and the eigenvector centrality value, which considers the centrality value of other related evaluation items among the centrality values, was estimated. The results of the eigenvector centrality analysis for the evaluation items are shown in Table VII. The evaluation item with the greatest eigenvector centrality is B (0.572), followed by D (0.550), E (0.393), C (0.355), and A (0.300) in that order appears. In the quantitative analysis results, the centrality values of evaluation items B and D were relatively high, and E, C, and A were low.

### TABLE VII. EIGENVECTOR CENTRALITY

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvector centrality value</td>
<td>0.300</td>
<td>0.572</td>
<td>0.355</td>
<td>0.550</td>
<td>0.393</td>
</tr>
</tbody>
</table>

Based on the analysis results, it can be divided into two groups according to the size of the centrality value. It can be divided into two groups: B, D, E, C, and A. When improving the evaluation system, it is appropriate to increase the score for the group with high relationship strength and to adjust the score for the group with low relationship strength.

The comprehensive improvement plan for the evaluation model, including the result of network centrality analysis, was determined by segmentation of evaluation item B. The reason for this decision is: First, it was derived as the most influential indicator in the model as a result of network centrality analysis, while being an object that needs improvement according to the results of regression analysis. Second, in the case of evaluation items B and D, the centrality value is similar, but the score of B is twice as high, so it is necessary to adjust the score of evaluation item B downward. Third, evaluation item B has the largest number of sub-indicators, so there is a limit to representing all the characteristics of sub-indicators. Fourth, in the case of improvement evaluation item C according to the results of regression analysis, it is appropriate to improve B evaluation item first because it has a low influence on other...
evaluation items and is clearly classified as a group with B evaluation item. Therefore, this paper subdivided evaluation item B into two evaluation items. B1 is 'Challenge and specificity of vision and operational goals', 'Excellence of Expected Performance and Utilization Plan', B2 is 'Justification and Necessity of Designation', 'Suitability of goals, project contents, research methods, etc.', 'Center It was grouped under 'fostering and operation support plan'. For the improved model, logistic regression analysis is re-executed to examine the validity of the improvement.

### D. Improvement Model Evaluation

The analysis results for the improvement model are shown in Tables VIII and IX. Through the Nagelkerke R2 index and the Hosmer-Lemeshow, a goodness-of-fit test of the model, the degree of explanation of the dependent variable by independent variables was about 33.6%, which was about 12% improved. As for the goodness of fit of the model, the value of the significance probability p was greater than 0.05, so the improved model also did not reject the null hypothesis.

![Table VIII. ANALYSIS RESULT OF HOSMER-LEMESEHOWTEST](image)

### VI. CONCLUSION

This paper identified problems in which the initial evaluation items and points assigned in the national supercomputer specialized center designation evaluation institution were consistently applied until the end of the project, and proposed a sustainable evaluation model improvement plan using evaluation result data. The difference from previous studies is that a network centrality analysis was newly performed to quantitatively analyze the strength of the relationship between evaluation items, and it was reflected in the improvement of evaluation items and indicators. As a result of the analysis, it was confirmed that the improvement plan of regrouping and subdividing the evaluation items using the eigenvector for the evaluation items was appropriate. It is expected that the results of this thesis will be used to continuously improve the designation institution and that excellent specialized centers will be selected.

The limitations of this thesis are that it has not been able to secure a lot of evaluation result data because many specialized centers have not yet been designated, and the effect of the improvement model has been proven only through statistical analysis. Therefore, in the 2023 second specialized center designation stage, plans are being established to apply the improved model in consultation with government agencies and re-verify the effect using the evaluation result data.

### ACKNOWLEDGMENT

This research was carried out through major project (K-23-L02-C03-S01) support of Korea Institute of Science and Technology Information.

### REFERENCES


