Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 270 (2025) 1727–1736

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

Proposal and Development of Tools for Visualizing the Relationship Between Childcare Records and Children's Qualities and Abilities Using Generative AI

Shirabe Yuji[a], Naoki Kato[b], Atsuo Hazeyama[b]

[a]*Graduate School of Education, Tokyo Gakugei University, Nukuikita-machi 4-1-1, Koganei-shi, Tokyo, Japan*
[b]*Tokyo Gakugei University, Nukuikita-machi 4-1-1, Koganei-shi, Tokyo, Japan*

Abstract

In the National Curriculum Standard for Kindergartens (revised in 2017), it is stated that kindergartens shall endeavor to nurture qualities and abilities. In order to support teachers in understanding children's development, this paper proposes tools that use generative AI to assess the degree of relevance (relatedness) between observed behaviors of children described in childcare records and qualities and abilities. Multiple methods based on different prompts were designed to calculate the relatedness score for each sentence, and validity was examined by comparing these scores to those assessed by experts. The results for a "foundation of knowledge and skills" and a "basic abilities of thinking, judgement, and expression, etc." showed high agreement with expert assessments under several methods. Although the overall agreement rate for the "learning dispositions and humanity, etc. to advance towards learning" was low, all methods showed correlation with expert assessments. Furthermore, three visualization methods were proposed, and received a certain level of positive evaluation by teachers and university students, suggesting that they could serve as helpful references for understanding children and grasping the qualities and abilities.

Keywords: Qualities and Abilities; Childcare Records; Generative AI

1. Introduction

In the National Curriculum Standard for Kindergartens (revised in 2017), it is stated that kindergartens shall endeavor to nurture qualities and abilities; a "foundation of knowledge and skills", a "basic abilities of thinking, judgement, and expression, etc.", and the "learning dispositions and humanity, etc. to advance towards learning". (hereafter referred to as the three elements) [1]. A key factor behind this emphasis is the growing global recognition

10.1016/j.procs.2025.09.293

of early childhood education, as seen in the OECD's focus on developing "key competencies" [2]. When formulating and implementing curricula, it is important to bear these qualities and abilities in mind and improve the quality of educational activities through a positive PDCA cycle based on an understanding of young children [3].

However, Hiraki has pointed out that while the 2017 revision newly introduced the concept of "qualities and abilities", the fact that "Goals and Content" in the 2017 revision of the National Curriculum Standard for Kindergartens remain almost unchanged from those in the 2008 revision has raised concerns about potential confusion in the field of early childhood education when developing instructional plans [4]. Furthermore, research by Yuji et al., in which childcare practitioners and pre-service teachers were asked to evaluate the presence of the three elements in video footage of childcare practice, has revealed that certain behaviors are easier to identify while others are more difficult to identify, and experience influences the ability to interpret children's behavior [5].

Although a clear grasp of the three elements is critical in understanding children, the difficulty of identifying them and the impact of practitioners' varying levels of experience point to the value of supporting teachers' capacity to detect the three elements in practice. Moreover, focusing on childcare records (Fig. 1), which contain a wealth of daily, practice-based descriptions, enables analysis directly tied to real-world settings, promotes reflection among practitioners, and may lead to higher-quality educational practices.

The present study proposes tools that use generative AI to visualize how closely each sentence describing children's behavior is related to the three elements from childcare records. By using these tools, teachers can grasp relatedness between children's behavior and the three elements. The main goal of the tools is to support teachers in understanding children and in formulating or enhancing instructional plans. Through utilization of this tool, a deeper understanding of the content of three elements may be achieved. Moreover, identifying children's behavior from the perspective of the three elements may enhance the quality of childcare records and strengthen educators' observational skills. In this study, only written childcare records are targeted and descriptions of not only the behavior of individual children but also that of the entire class are included as text data.

Previous research on extracting qualities and abilities from text data includes methods by Hagiwara et al. that use supervised machine learning and morphological analysis to visualize the development of qualities and abilities from reflective learning texts [6]. In addition, Fareri et al. developed SkillNER, a supervised NER system trained on a scientific corpus, which can automatically retrieve soft skills from text [7]. Somers et al. proposed free-text validity ensemble using NLP that can assess students' justification and confidence from their responses to short questions [8].

However, there appear to be no prior studies that develop methods using generative AI to specifically identify or evaluate the three elements, nor any that develop a system for recognizing these abilities in early childhood.

Validity is examined by comparing the method's results to human expert assessments. Additionally, questionnaires administered to in-service kindergarten teachers and university students clarify the usefulness of the proposed method of visualizing the analytical results.

June 14 (Thu)

The children formed a new living-group and decided they wanted an animal-themed name.

H, N, T, R, and S began tossing around ideas.

N, R: "How about Hedgehogs? They're cute."

S: "Hmm, I like Rabbits, but Hedgehogs are cute too, so that's fine."

T: "Hedgehog is fine with me"

S: "But hedgehogs stick out their spines and that might hurt. I'm not sure I like that. Teacher, do hedgehogs really shoot out their spines?"

Fig. 1. Examples of sentences in childcare records (author's translation and adaptation from [3])

2. Study of the Method for Reading Qualities and Abilities

2.1. Basic Concept

This section proposes and designs a method to determine, at the sentence level, how closely the behavior of children in childcare records is connected to the three elements. Analysis is conducted on sentence by sentence to allow educators to notice critical parts of childcare record texts by extracting sentences with high relatedness.

Because the National Curriculum Standard for Kindergartens and related guidelines offer limited detail on the three elements, creating an exhaustive list of all potential behaviors that indicate the three elements is challenging. Simple keyword-based judgements may fail to capture the nuances of relatedness accurately. Therefore, this study leverages a large-scale generative AI, which has learned from extensive textual data and assigns relatedness scores based on contextual understanding.

When designing the proposed method, multiple approaches were developed by varying the information and level of specificity in the prompts (instructions) given to the AI.

2.2. Rubric as the Basis for Relevance

A rubric was developed to standardize the criteria for determining relatedness scores and to guide both the AI-based calculation and experts' manual evaluations (Table 1). Two experts specializing in early childhood education participated in designing this rubric.

The National Curriculum Standard for Kindergartens indicates about evaluation that "it is not something that can be accomplished by comparing the child to other children or by rating an achievement against certain criteria" [1]. Accordingly, this rubric categorizes sentences into one of five relatedness levels (from 1 to 5) based on whether the child's behavior indicating each element is directly or indirectly discernible, or whether it can be definitively concluded that the child is not exhibiting that element.

Table 1. A rubric standardizes the criteria for determining relatedness scores

Relatedness	The criteria for determining scores
5	The sentence explicitly shows the child exhibiting the ability, and it can also be clearly inferred from the child's behavior that the ability is indeed being demonstrated.
4	The sentence does not explicitly show the child exhibiting the ability, yet it can be clearly inferred from the child's behavior.
3	The sentence does not explicitly show the child exhibiting the ability, but there is a possibility of it being demonstrated, based on the child's behavior.
2	The sentence does not explicitly show the child exhibiting the ability, and the child's behavior suggests that the ability is not being demonstrated, although this cannot be definitively concluded.
1	The sentence does not explicitly show the child exhibiting the ability, and from the child's behavior it can be definitively concluded that the ability is not present.

2.3. Detailed Design of the Method for Reading Qualities and Abilities

Building on the rubric described in Section 2.2, the proposed method determines how closely each sentence describing a child's behavior in the childcare record relates to the three elements, using an integer from 1 to 5. This section explains how prompts for the generative AI were designed to calculate relatedness.

To investigate optimal prompting, a "basic prompt" and a "simplified prompt" (created by simplifying the basic prompt) were created, as well as a "control prompt" that either incorporates or omits additional information.

In creating the simplified prompt, some of the 26 principles designed by Bsharat et al. to improve and enhance the quality of responses of large language models [9] were adopted. Bsharat et al. found applying these principles improves model performance such as LLaMA-1/2, GPT-3.5/ GPT-4.

2.3.1 Basic Prompt and Simplified Prompt

Fig. 2 shows the basic prompt. To clarify the AI's role, the phrase "You are an expert in early childhood education" is included. The rubric (Section 3.2) is also provided in the prompt.

For comparative purposes, a simplified prompt (Fig. 3) was created by reducing the level of detail in the instructions. Adhering to one of Bsharat et al.'s 26 principles: "Employ affirmative directives such as 'do,' while steering clear of negative language like 'don't'" [9], and certain expressions in the rubric were abbreviated, leading to an overall reduction of 86 Japanese characters.

Please analyze the following childcare-record sentences and evaluate each one on a scale of 1 to 5, indicating how strongly it relates to qualities and abilities (a “foundation of knowledge and skills”, a “basic abilities of thinking, judgement, and expression, etc.” and the “learning dispositions and humanity, etc. to advance towards learning”).
The focus of the evaluation is not on which qualities or abilities should be developed in the child, but rather on which qualities or abilities are currently being demonstrated by the child’s behavior.

5 The sentence explicitly shows the child exhibiting the ability, and it can also be clearly inferred from the child’s behavior that the ability is indeed being demonstrated.
4 The sentence does not explicitly show the child exhibiting the ability, yet it can be clearly inferred from the child’s behavior.
3 The sentence does not explicitly show the child exhibiting the ability, but there is a possibility of it being demonstrated, based on the child’s behavior.
2 The sentence does not explicitly show the child exhibiting the ability, and the child’s behavior suggests that the ability is not being demonstrated, although this cannot be definitively concluded.
1 The sentence does not explicitly show the child exhibiting the ability, and from the child’s behavior it can be definitively concluded that the ability is not present.

Fig. 2. Basic prompt

Please analyze the following childcare-record sentences and evaluate each one on a scale of 1 to 5, indicating how strongly it relates to qualities and abilities (a “foundation of knowledge and skills”, a “basic abilities of thinking, judgement, and expression, etc.” and the “learning dispositions and humanity, etc. to advance towards learning”).
The focus of the evaluation is on which qualities or abilities are currently being demonstrated by the child’s behavior.

5 The sentence explicitly shows the child exhibiting the ability, and it can also be clearly inferred that the ability is indeed being demonstrated.
4 The sentence does not explicitly show the child exhibiting the ability, yet it can be clearly inferred.
3 The sentence does not explicitly show the child exhibiting the ability, but there is a possibility of it being demonstrated.
2 The sentence does not explicitly show the child exhibiting the ability, and it is suggested that the ability is not being demonstrated, although this cannot be definitively concluded.
1 The sentence does not explicitly show the child exhibiting the ability, and it can be definitively concluded that the ability is not present.

Fig. 3. Simplified prompt

2.3.1. Control Prompts

2.3.1.1. Providing Explanations of the Three Elements

A control prompt was devised to examine whether including descriptive explanations of the three elements from the National Curriculum Standard for Kindergartens affects relatedness scoring. Specifically, statements from official guidelines explaining the three elements were inserted (Fig. 4). The basic prompt text was then modified to indicate that the AI should “refer to the information based on the National Curriculum Standard for Kindergartens.”

In order to foster a foundation to embrace a zest for living, kindergartens shall endeavor to nurture all of the following qualities and abilities.

(1) A "foundation of knowledge and skills" that enables one to feel, realize, understand, and become able to do things through rich experiences

(2) A "Basic abilities of thinking, judgement, and expression, etc." that enables one to think, try, devise, and express using what one realizes or becomes capable of doing

(3) The "Learning dispositions and humanity, etc. to advance towards learning" that endeavors to maintain a better life while emotions, motivation and attitudes develop

Fig. 4. Explanations of the three elements

2.3.1.2. Providing Concrete Examples of the Three Elements

In the National Curriculum Standard for Kindergartens, “Ideal Image by the End of Childhood”, a specific form envisioned for the completion of kindergarten by children whose qualities and abilities have been developed is explicitly defined. To verify whether providing such examples influences relatedness scoring, relevant excerpts for each element were taken from these sections. For instance, Fig. 5 shows the examples used for a “foundation of knowledge and skills”.

Feel, realize, understand, and become able to do things through rich experiences
Use one's mind and body
Be aware of what needs to be done
Share ideas and thoughts with each other
Come to understand right from wrong
etc.

Fig. 5. The examples used for a "foundation of knowledge and skills"

2.4. Generative AI and Parameter Settings

The GPT-4o API from OpenAI was adopted as the generative AI model [10]. To ensure consistent relevance assessments the temperature parameter was set to 0.2, and the maximum number of tokens was limited to 1500.

3. Validation of Relevance Calculation

This section presents a comparison between the relatedness scores produced by the proposed methods and the scores assigned by six experts in order to assess the validity of the AI-based scoring. The experts included three university faculty members specializing in early childhood education and three kindergarten teachers with over ten years of childcare experience.

3.1. Overview

3.1.1. Targeted Examples

The text passages for validation were selected from practice examples in the publication "Designing an Environment That Resonates with Children in Centers for Early Childhood Education and Care" (2022) [11]. Of the 13 examples described in Chapter 3, 6 examples were randomly chosen. Three sentences describing children's behavior were further chosen at random, resulting in a total of 18 sentences. Sentences containing actions by multiple different subjects were excluded from the selection (e.g., "Child A did X, while Child B did Y").

3.1.2. Method for Relevance Evaluation and Calculation

The six experts were provided with the full text of each of the six examples. After reading the entire text, they determined the relatedness score for each of the 18 selected sentences using the rubric (Section 2.2).

For the AI-based scoring, only the 18 sentences were input into the system. To account for variability in AI outputs, 10 trials were conducted, and the average of those 10 trials was used in subsequent analyses (Section 3.2).

3.1.3. Methods to Compare

Eight different scoring methods (Methods a to h) were created by combining either the basic prompt or the simplified prompt with the presence or absence of two additional pieces of information (the explanatory statements and concrete examples for the three elements). Table 2 summarizes the differences among these methods.

3.2. Results

This section presents the analysis and discussion of the results of the variance analysis of the relevance among the methods, the agreement rates with expert evaluations, and correlation analyses.

3.2.1. ANOVA

A one-way Welch's ANOVA ($\alpha = .05$) was performed on the average relatedness scores (10 trials) of the 18 sentences under the eight methods. The analysis showed a significant difference ($F = 5.33$, $p < .001$, CI [3.075, 10.972], $\eta^2 = 0.215$, $F(7, 136, 0.05) = 2.08$) among the eight methods for a "foundation of knowledge and skills". However, no significant differences were observed for a "basic abilities of thinking, judgement, and expression, etc."

Table 2. The differences among methods

Methods	Level of Specificity	Additional Information	
		Explanations	Examples
a	Basic	○	×
b		×	○
c		○	○
d		×	×
e	Simplified	○	×
f		×	○
g		○	○
h		×	×

($F(7, 136) = 0.49$, $p = .84$, CI [0.378, 3.408], $\eta^2 = 0.025$) or the "learning dispositions and humanity, etc. to advance towards learning" ($F(7, 136) = 0.32$, $p = .95$, CI [0.324, 3.072], $\eta^2 = 0.016$).

3.2.2. Agreement Rate

Among the six experts, the relatedness evaluations for each element were considered in agreement if four or more experts (i.e., a majority) assigned the same score to a sentence. Based on this criterion, there was consensus on 10 sentences for a "foundation of knowledge and skills", 9 sentences for a "basic abilities of thinking, judgement, and expression, etc.", and 7 sentences for the "learning dispositions and humanity, etc. to advance towards learning" Table 3 shows the percentage of sentences in which each methods derived relatedness score (averaged over 10 trials) matched the majority expert rating.

For a "foundation of knowledge and skills", three methods achieved an agreement rate of 70% or higher, with the highest agreement of 87.0% observed in Method g, which utilized the simplified prompt along with explanations and concrete examples of the three elements.

For a "basic abilities of thinking, judgement, and expression, etc.", three methods reached an 80% agreement rate, and all methods showed at least 50% agreement. Overall, agreement rates were relatively high.

In contrast, for the "learning dispositions and humanity, etc. to advance towards learning" all methods had agreement rates below 50%. The highest rate was 47.1% in Method e, indicating that overall consistency was lower for this element.

Table 3. Agreement rate

Qualities and Abilities	a	b	c	d	e	f	g	h
Foundation of knowledge and skills（n＝100）	38.0%	41.0%	63.0%	31.0%	53.0%	82.0%	87.0%	70.0%
Basic abilities of thinking, judgement, and expression, etc.（n＝90）	65.6%	80.0%	64.4%	80.0%	80.0%	66.7%	53.3%	77.8%
Learning dispositions and humanity, etc. to advance towards learning（n＝70）	30.0%	24.3%	31.4%	12.9%	47.1%	30.0%	42.9%	38.6%

3.2.3. Correlation Analysis

An analysis of correlations between the average AI-based relatedness scores (10 trials) and the average expert-relatedness scores was performed using data for 18 sentences. Fig. 6 shows the correlation coefficients and confidence intervals.

For a "foundation of knowledge and skills", all methods exhibited moderate to strong positive correlations with relatively narrow confidence intervals, suggesting consistent scoring trends. The highest correlation ($r = 0.69$) occurred with Method h, which used the simplified prompt without additional explanatory information.

For a "basic abilities of thinking, judgement, and expression, etc.", seven of the eight methods (excluding Method g) showed moderately to strongly positive correlations. Method b had the highest correlation ($r = 0.79$).

For the "learning dispositions and humanity, etc. to advance towards learning" correlation coefficients were large for all methods, with relatively narrow confidence intervals, indicating strong positive correlations. The highest correlation ($r = 0.77$) was observed with Method d.

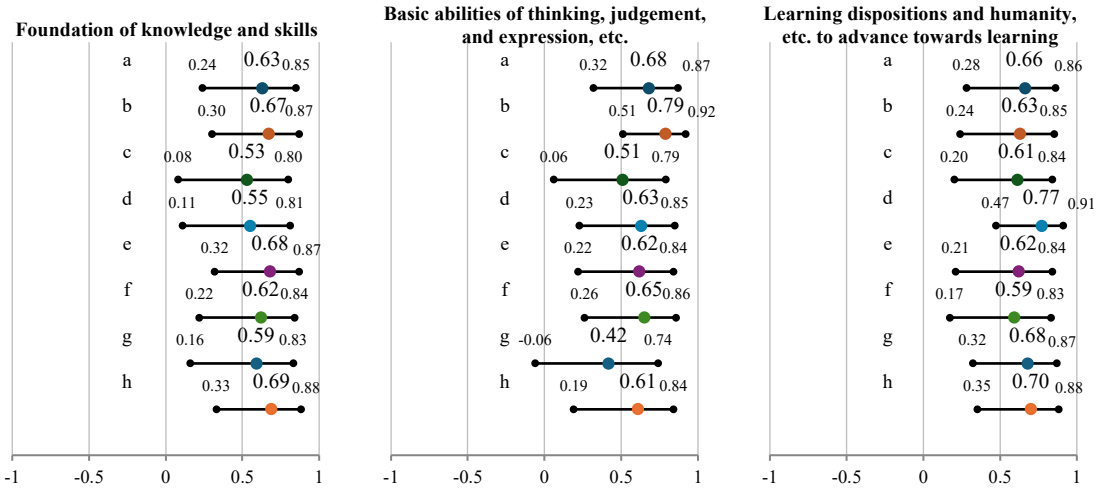

Fig. 6. Correlation coefficients and confidence intervals

3.3. Discussion

3.3.1. A "foundation of knowledge and skills"

Based on the above results, for a "foundation of knowledge and skills", method g (using the simplified prompt plus the explanation and concrete examples of the three elements) showed the highest agreement rate with the experts at 87.0%. Likewise, method c (which uses the basic prompt plus the same explanations and examples) had the highest agreement rate (63.0%) among the basic-prompt-based methods. These findings suggest that providing both an explanation and concrete examples of the three elements may positively affect relevance calculations in this area.

Comparing the methods that use the basic prompt (a, b, c, d) with those that use the simplified prompt (e, f, g, h), we see that when the additional information was the same, the simplified prompt methods yielded agreement rates 15–41 points higher than the basic prompt methods. This suggests that simplifying the prompt may improve the calculation of relevance.

3.3.2. A "basic abilities of thinking, judgement, and expression, etc."

For a "basic abilities of thinking, judgement, and expression, etc.", overall agreement rates were high, and 7 of the 8 methods (all except g) showed moderate to strong correlations. Meanwhile, method g, which had the lowest agreement rate among the 8 methods, did not show a strong correlation. Moreover, even among only the basic-prompt methods a, b, c, and d, the methods that provided both explanations and examples of the three elements showed the lowest agreement rates. Consequently, although explanations and examples of the three elements appeared beneficial for a "foundation of knowledge and skills", they may have reduced agreement for a "basic abilities of thinking, judgement, and expression, etc.".

3.3.3. The "learning dispositions and humanity, etc. to advance towards learning"

For the "learning dispositions and humanity, etc. to advance towards learning" all methods showed agreement rates below 50%. The highest (47.1%) was Method e, which combined the simplified prompt with explanations of the three elements. One possible reason for these lower agreement rates compared to other elements is that the "learning dispositions and humanity, etc. to advance towards learning" involves children's internal states, such as feelings and motivation, which are inherently more difficult to observe. Indeed, Yuji et al. noted that the "learning dispositions and humanity, etc. to advance towards learning" is more challenging to identify, whereas a "basic abilities of thinking, judgement, and expression, etc." is relatively easier to identify [5].

Although overall agreement was low, all methods showed notable correlation with expert evaluations, suggesting a consistent scoring trend by GPT-4o that is at least partially aligned with experts' tendencies.

When comparing the basic-prompt methods (a–d) and simplified-prompt methods (e–h) that include the same additional content, the simplified prompts showed an agreement rate improvement of between 5.7 and 25.7 points. This again suggests the potential advantage of simplifying prompts. Future work may focus on refining prompts to improve the agreement rate for this element.

4. Study of Visualization Methods

This section proposes three methods for visualizing analysis results derived from the approaches presented in Section 2 and clarifies their usefulness based on questionnaires administered to both teachers and university students.

4.1. Proposed Visualization Methods

Three visualization methods were designed for presenting the analysis of childcare records. Each combines text output (referred to below as "analysis text"), which includes the relatedness scores and reasons for those scores, in one of the following three formats:

- A: Displaying only the analysis text (B with the scores removed)
- B: Displaying both the analysis text and the relatedness scores (The left of Fig. 7)
- C: Underlining sentences with high relatedness in different colors corresponding to each of the three elements, coupled with radar charts (The right of Fig. 7)

Although the five-point rubric was used for the methods in Section 2, at the time of prototyping the visualization methods, the system was instructed to provide scores between 0 and 1. Methods A and B analyze one entire example at a time, displaying the overall relatedness scores for that example. The analysis text was generated by instructing Method a in 3.1.1. to explain the reasoning of scoring. Method C processes text one sentence at a time, uses color-coded underlines to highlight sentences with high scores, and shows radar charts when hovering over each sentence.

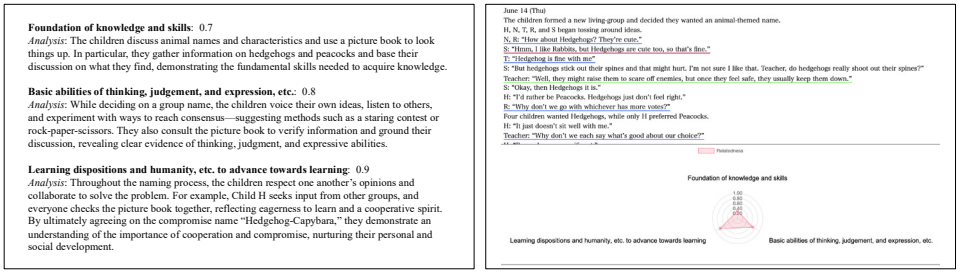

Fig. 7. Visualization method B (The analysis text and relatedness scores) and C (Underlining and radar charts)

4.2. Evaluation of the Visualization Methods

This section demonstrates the usefulness of the proposed visualization methods based on the results of questionnaires administered to kindergarten and other early childhood teachers (48 respondents with 2–36 years of experience) and to university students (94 respondents aspiring to be an educator).

4.2.1. Target Example

The childcare record selected for visualization was "Thinking of Group Names Together" (for five-year-olds), from "Creating Instructional Plans and Developing Childcare Connecting Children's Thoughts" (2021) [3].

4.2.2. Questionnaire Items

Two questions were asked about each of the three proposed visualization methods (A, B, C). Participants responded using a five-point scale, a range from 2 ("Strongly agree") to −2 ("Strongly disagree"):

1. "Are the displayed items helpful for understanding children?"
2. "Are the displayed items useful for identifying qualities and abilities?"

The first question asks whether the visualization methods help in understanding children, while the second asks whether they help in understanding the content of the three elements.

Participants were then asked which of the three visualization methods they would choose for use in real childcare settings, and were given the opportunity to provide free-text feedback on each visualization method.

4.3. Questionnaire Results and Discussion

4.3.1. Results and Discussion of the Teacher Survey

Fig. 8 shows the responses to questions 1 and 2. The numbers in the figure represent the mean and standard deviation of the teachers' ratings. All three methods received positive evaluations suggesting they could facilitate understanding children and recognizing their qualities and abilities. Method C earned the highest ratings on both questions.

Regarding the question about which method teachers would choose in practice, 35.4% (17 teachers) selected Method A, 12.5% (6 teachers) selected Method B, and 52.1% (25 teachers) selected Method C.

In the free-response section, teachers commented on Visualization Method C, for example: "It is easy to understand even for teachers who have fewer years of experience, and it can serve as a guide when taking records".

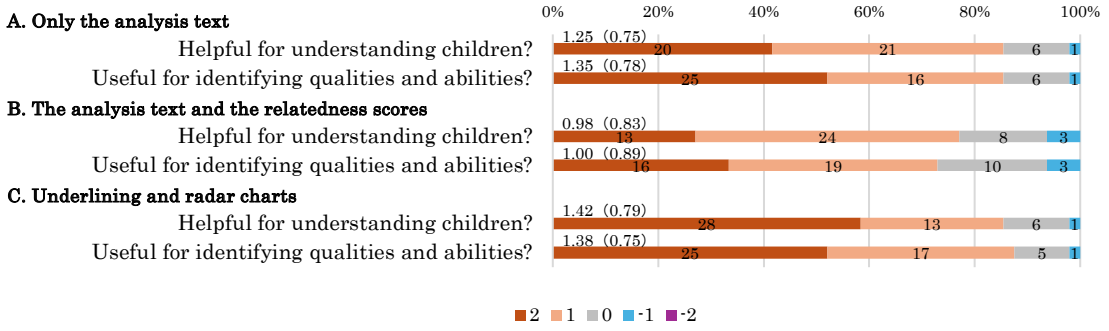

Fig. 8. The responses to questions about visualization method and average score (standard deviation)

4.3.2. Results and Discussion of the Student Survey

Fig. 9 shows the responses to questions 1 and 2 for the students at the teacher-training university. Again, Method C received the highest average ratings on both questions. At the same time, all three methods received reasonably positive evaluations.

For the question "Which method would you choose in a childcare setting?", 27.7% (26 students) selected Method A, 22.3% (21 students) selected Method B, and 50.0% (47 students) selected Method C.

Free responses about Method C included positive statements, such as: "The visual differences are easy to understand", However, there were also comments like: "It's difficult to understand because it's just underlining".

5. Conclusion

In aiming to support teachers' understanding of children, this study proposed tools that use generative AI to analyze childcare records and estimate the degree to which children's behavior is related to qualities and abilities defined in the National Curriculum Standard for Kindergartens.

Multiple approaches based on different prompts were developed to calculate sentence-level relatedness, and validity was evaluated through comparisons with expert assessments. The results for a "foundation of knowledge and skills" showed high agreement in several methods, suggesting that explanatory content about the three elements, concrete examples, and simpler instructions may have beneficial effects. For a "basic abilities of thinking, judgement, and expression, etc.", overall agreement rates were high, with moderate to strong correlations observed in seven methods. Although agreement rates for the "learning dispositions and humanity, etc. to advance towards

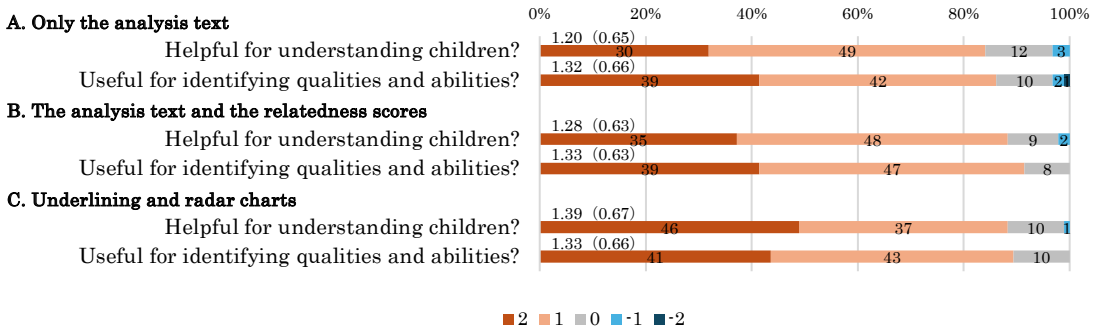

Fig. 9. The responses to questions about visualization method and average score (standard deviation)

learning" were generally lower, all methods demonstrated correlation with expert assessments, suggesting that GPT-4o-based scoring may align with experts' tendencies.

Furthermore, three visualization methods were proposed for presenting the analytic results. Questionnaires completed by in-service teachers and university students confirmed the usefulness of all methods. Notably, the method employing underlines and radar charts (Method C) received the highest evaluations from both groups.

Future task is to develop practical tools incorporating both the scoring and visualization methods proposed here.

Acknowledgements

Sincere gratitude is extended to the experts who assisted in developing the rubric and validating the relatedness-scoring methods, as well as the teachers and university students who generously participated in the questionnaires.

Declaration of generative AI and AI-assisted technologies in the writing process

OpenAI o1 and o3 were used to support translation tasks in preparing this manuscript. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the final published content.

References

[1] Ministry of Education, Culture, Sports, Science and Technology (2017) "The National Curriculum Standard for kindergartens" https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2019/10/11/1401777_002.pdf (accessed 2025-03-31)

[2] OECD (2005) "The Definition and Selection of key Competencies: Executive Summary"

[3] Ministry of Education, Culture, Sports, Science and Technology (2021) "Creating Instructional Plans and Developing Childcare Connecting Children's Thoughts" *CHILD HONSHA Co.,Ltd.*

[4] Hiraki, Hitoshi (2019) "Consideration of "Aim and Contents" Based on "Qualities and Abilities that are Aimed to Develop" in Early Childhood, *Kanazawa Seiryo University Human Sciences* 12 (2): 75-80.

[5] Yuji, Hiroki, Takaoka, Yukiko and Yuji,Yumi. (2022)"The Study of Valuation of the "Ideal Image by the End of Childhood," "Qualities and Abilities," "Proactive and Dialogic Deep Learning," and the "Five Areas" in the National Curriculum Standard for Kindergartens", *Research Bulletin of Naruto University of Education* 37: 313-326

[6] Hagiwara, Kohei, Maruyama, Kohei and Morimoto, Yasuhiko. (2023) "Development of an Extracting Method of a Description Related to Fostering of Students' Competencies from Text Data of Reflections on Learning by Supervised Learning and Morphological Analysis" *Japan Journal of Educational Technology* 47: Suppl. 241-244

[7] Fareri, Silvia et al. (2021) "SkillNER: Mining and mapping soft skills from any text" *Expert Systems With Applications* 184: 1-10

[8] Somers, Rick, Cunningham-Nelson, Samuel and Boles, Wageeh. (2021) "Applying natural language processing to automatically assess student conceptual understanding from textual responses" *Australasian Journal of Educational Technology* 37(5): 98-115

[9] Bsharat, Sondos Mahmoud, Myrzakhan, Aidar and Shen, Zhiqiang. (2023) "Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4" https://arxiv.org/pdf/2312.16171 (accessed 2025-03-31)

[10] OpenAI "Hello GPT-4o" https://openai.com/index/hello-gpt-4o/ (accessed 2025-03-31)

[11] Cabinet Office, Government of Japan, Ministry of Education, Culture, Sports, Science and Technology and Ministry of Health, Labour and Welfare. (2022) "Designing an Environment That Resonates with Children in Centers for Early Childhood Education and Care" *Froebel-kan Co.,Ltd.*