Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1, 18-26
2026
Publisher: Learning Gate
DOI: 10.55214/2641-0249.v8i2.11818

Development of the gFUMIE test for measuring implicit gender stereotypes and a preliminary validation with undergraduate students using doctor and florist

Shin AKITA[1], Hideko MORI[2], Kazuo MORI[3*]
[1]Matsumoto University, Japan; shin.akita@t.matsu.ac.jp (S.A.).
[2]Bunka Gakuen University Nagano College, Japan; mori-h@gakuen-nagano.ac.jp (H.M.).
[3]Tokyo University of Agriculture and Technology, Japan; kaz-mori@cc.tuat.ac.jp (K.M.).

Abstract: We developed a new paper-and-pencil method, the gender-FUMIE (gFUMIE) test, for assessing gender stereotypes in children by adapting the FUMIE test (Mori, Uchida, & Imada, 2008). The test comprises a series of gender-classification tasks using kinship terms, such as "mother" and "father," with a target concept randomly interspersed among them. By measuring differences in how quickly the target is classified as male or female, gender stereotypes can be assessed efficiently. Before applying the test to children, a preliminary validation was conducted with 70 undergraduate students (41 males and 29 females) using two typical gender-stereotyped occupations: "doctor = male" and "florist = female." The results revealed a distinct feminine stereotype for "florist," whereas no clear gender stereotype was observed for "doctor." Although further verification using other occupations is necessary, the findings suggest that the assessment principle of the gFUMIE test is effective for detecting gender stereotypes using a child-friendly procedure. We expect that a children's version of the gFUMIE test will serve as a new method for studying the development of gender stereotypes.

Keywords: *Elementary school children, Implicit measurement, Japanese students, Paper-and-pencil IAT.*

1. Background

While gender equality has gained social attention, studies show that individuals endorsing egalitarian values may still unconsciously hold stereotypical gender roles. The Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) is a common measure of implicit gender bias. However, its reliance on a computer environment limits its use in school education and large-scale surveys.

To address this limitation, Mori et al. (2008) modified the IAT into a paper-and-pencil format and developed the FUMIE test, which can be administered efficiently to groups within a short period. The FUMIE test offers a simple method for measuring implicit associations without computers and has been widely used by elementary school students to university students. For example, Akita and Mori (2022) used a version adapted for upper elementary school children (Akita, Tsushima, Saito, & Mori, 2019) in a sixth-grade social studies lesson on gender equality, which improved boys' implicit images of women. Additionally, Akita and Mori (2023) administered the same FUMIE test three years later, confirming that the improvement was maintained.

Mori et al. (2019) revised the FUMIE test for third- and fourth-grade elementary students. Consequently, Akita and Mori (2025) used this version to measure the effect of a gender equality lesson for fourth graders. They found no significant change in students' implicit images of women before and after the lesson. This suggests that the FUMIE test may not be effective for children below third- or fourth-grade level. Nonetheless, to study the development of gender stereotypes, a new measurement tool applicable to lower-grade elementary students is necessary.

* Correspondence: kaz-mori@cc.tuat.ac.jp

Whitford (2023) conducted a qualitative study with second-grade children in the United States, finding that although they spoke positively about gender equality, they still held gender-stereotypical thoughts implicitly. This suggests that a discrepancy between explicit and implicit awareness emerges early in development, highlighting the importance of measuring implicit attitudes in young children.

To facilitate research, Thomas, Burton Smith, and Ball (2007) adapted the IAT for children aged three to seven. Similarly, Greenwald, the IAT's original developer, and colleagues created a version suitable for four-year-olds (Cvencek, Greenwald, & Meltzoff, 2011), finding that "flowers" were implicitly preferred over "bugs." These adaptations used pictures on screens, enabling preliterate children to perform the tasks. However, despite these innovations, the IAT still requires a computer environment, which limits its ease of administration in typical school settings.

We aimed to develop the gender-FUMIE (gFUMIE) test, an improved version of the FUMIE test that remains easy to administer in classroom settings while enabling more convenient measurement of implicit gender attitudes. As a preliminary step, this study constructed a student version of the gFUMIE test and examined its validity. Before developing a children's version, we sought to confirm whether the student version could detect theoretically predicted gender–occupation stereotypes, specifically, "Doctor = Male" and "Florist = Female", among university students.

2. Research Purposes

The purpose of this study was to develop the gender-FUMIE (gFUMIE) test, a paper-and-pencil instrument designed to detect gender–occupation stereotypes among lower-grade elementary school children. To verify its validity, a student version of the gFUMIE test was constructed and administered to examine whether it could detect gender stereotypes related to the occupations "doctor" and "florist."

3. Construction of the gFUMIE Test for Lower-Grade Elementary School Children

3.1. Principles of the FUMIE Test and Its Merits

The gender–occupation stereotype detection test developed in this study was based on the FUMIE test, maintaining its core principle while making it suitable for lower-grade elementary school children. The fundamental principle of the FUMIE test involves two components: a basic classification task that requires sorting words with positive or negative meanings, and a cognitive conflict task that compels participants to classify a target word as either "good" or "bad" to measure its implicit image.

In the basic task, words with positive and negative meanings are used. Participants are asked to judge whether each word has a "good" or "bad" meaning and to mark it with a circle ("O") or a cross ("X"), respectively. The words are carefully selected so that their valence is clearly positive or negative in advance. Consequently, participants can perform this classification task almost automatically.

In the cognitive conflict task, a "target word" (e.g., "woman") is embedded among other words, and participants are instructed to classify the target as either "good" or "bad." If a participant holds a positive image of "woman," classifying it as "bad" induces cognitive conflict, resulting in slightly slower responses than when classifying "woman" as "good." Conversely, if a participant holds a negative image of "woman," classifying it as "good" causes cognitive conflict, leading to longer response times.

In other words, if a participant classifies a target word as "good" more quickly than as "bad," it suggests a positive implicit image of that word. Conversely, faster classification as "bad" indicates a negative implicit image.

In summary, when participants classify a target word as either "good" or "bad," those who classify it more quickly as "good" are considered to hold a positive implicit image of that word. Conversely, if they classify it more quickly as "bad," it can be inferred that they hold a negative implicit image of it.

The FUMIE test involves alternating tasks of classifying words as "good" or "bad" for each row. The number of words marked with "O" or "X" within 20 seconds measures the participant's "classification speed." This method replaces traditional reaction-time measurement with work quantity within a fixed time. Its advantage is that the test can be administered to many participants simultaneously. In

Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

educational settings, the FUMIE test can be efficiently administered to an entire class by distributing answer sheets and conducting several 20-second sessions.

3.2. Revisions by Utilizing Illustrations and Modifying Response Formats

In this study, we aimed to develop a measurement method that maintains the fundamental principles and simplicity of the FUMIE test while allowing direct assessment of gender bias. The task replaced the "good–bad" classification with a "male–female" classification. Using familiar family-role words such as "mother," "father," "grandmother," and "grandfather," the task can be performed even by lower-grade elementary school children.

Since the task does not involve evaluating goodness or badness, participants mark a circle on either the left or right side of each word, where illustrations representing "male" and "female" are displayed. To make the distinction intuitive, response areas are color-coded, light blue for "male" and pink for "female."

Unlike the original FUMIE test, where "O" or "X" marks are written directly next to horizontally arranged words, the new version requires circling on the left or right side. Consequently, in this revised test, the words are arranged vertically rather than horizontally on the answer sheet.

The gender-stereotype task related to the cognitive conflict task of the FUMIE test involved presenting occupations like "doctor" and "florist" written in hiragana. Instructions, delivered via PowerPoint slides with illustrations of a "male doctor" or "female doctor" (or a "male florist" or "female florist"), directed children to classify each word as "male" or "female." For children who believe "florists are women," classifying "florist" as "male" creates cognitive conflict. This conflict is expected to increase response time and decrease the number of responses, reflecting the influence of gender stereotypes on task performance.

With these modifications, we developed a prototype to measure occupational gender stereotypes among lower-grade children. To enhance approachability, the word "test" was avoided in the title, and it was renamed The Boy–Girl Game (see Appendix A).

4. Validation of the gFUMIE Test with University Students

Would the gFUMIE test for lower-grade elementary school children be effective in detecting gender stereotypes as expected? To answer this, it should be administered directly to younger children. However, a preliminary validation with university students is useful to confirm whether the basic measurement principle of the gFUMIE test functions properly before testing younger children.

Mori, Akita, and Mori (2025) previously administered the adult version of the gFUMIE test to 44 students at a nursery education college and reported that it successfully detected the gender stereotype of "Florist = Female." In contrast, the stereotype of "Doctor = Male" was not detected. Only five male students participated, as the college's student body was predominantly female. The absence of the "Doctor = Male" stereotype may have been due to the gender imbalance among participants.

In this study, we reexamined education-major students with a balanced gender ratio to determine if stereotypes like "Florist = Female" and "Doctor = Male" could be identified. We also explored whether these gender stereotypes differed between male and female students.

4.1. Participants

Seventy university students (41 males and 29 females; mean age = 21.3 years for males and 21.2 years for females) participated in the study. The research's purpose and procedures, including the voluntary nature of participation, were fully explained. Submission of the completed answer sheet indicated informed consent. Participants were not asked to provide their names and were instructed to record only their age and gender.

4.2. Construction and Administration Procedure of the gFUMIE Test for Students

The layout of the test sheet follows the format of the gFUMIE test designed for lower-grade children. The student version was created using an Excel worksheet and arranged for horizontal printing on

Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

standard paper (A4 or B4 size). Thirteen vertical columns were prepared, each containing 50 words related to gendered family-role terms (e.g., father, mother, brother, sister) and occupational terms to be measured. Response cells were placed on both sides of each column, blue boxes on the left for "male" and pink boxes on the right for "female." Illustrations of male and female faces were printed at the top to clearly indicate the two gender categories. The test sheet contained 14 columns in total, but the final column was left unused to prevent end-of-task effort effects (see Appendix B).

The basic classification task required participants to identify the gender of family-role words and classify them quickly. As practice, the first column contained only these words, and participants performed the task using that column.

Gender stereotype measurement task involving six columns (second to seventh) where the occupation word "doctor" was randomly inserted among family-role words. In the following six columns (eighth to thirteenth), the word "florist" was similarly inserted. Each column featured a PowerPoint slide showing an illustration, prompting participants to classify the occupation as either "female doctor" or "male doctor," and similarly for "female florist" or "male florist." Participants were instructed to classify these words based on gender, consistent with their classification of family-role words.

The administration procedure involved presenting all test phases, instructions, practice, doctor task, florist task, and debriefing, visually via PowerPoint slides, with verbal directions from the experimenter. Each task, including practice, lasted 20 seconds, controlled by PowerPoint's automatic timing function.

After completing all tasks, a debriefing slide expressing appreciation was shown, and participants were asked to fill in the gender and age sections on the answer sheet. They were informed they could omit their age if they preferred not to disclose it. Submission of the answer sheet was regarded as consent to participate, and the sheets were collected. The entire procedure was conducted simultaneously in a classroom setting and took approximately 10 minutes, including the explanation period.

5. Results

Voluntary participation: All 70 students consented to participate and submitted their answer sheets.

Measurement index: For both "doctor" and "florist," the number of marked responses was counted separately for male and female classification tasks. The total for three rows was then converted to work performed per minute (60 seconds). Based on these values, the Implicit Gender Bias (IGB) score was calculated using the following formula. The IGB_{100} score indicates the difference between male and female conditions, standardized to classify 100 words. A positive IGB_{100} value signifies a bias toward males for the occupation, while a negative value indicates a bias toward females.

$$IGB_{100} = \frac{(\text{Number of Words Done in Male Task} - \text{Number of Words Done in Female Task})}{(\text{Number of Words Done in Male Task} + \text{Number of Words Done in Female Task})} \times 100$$

Exclusion of outliers: According to the FUMIE Test Administration Manual (Uchida & Mori, 2018), data from participants with IGB_{100} scores deviating more than ±2 SD from the mean were excluded, five male and one female student. Consequently, 36 male and 28 female students' data were included in the analyses.

Analysis of IGB scores: The mean IGB_{100} scores for both "doctor" and "florist" were calculated separately for male and female students. Contrary to initial predictions, both male and female students classified the word "doctor" slightly faster in the female condition (male students: $IGB_{100} = -1.53$; female students: $IGB_{100} = -0.47$). However, the 95% confidence intervals ($-3.22 <$ male $IGB_{100} < 0.17$; $-1.74 <$ female $IGB_{100} < 0.79$) included zero in both cases, indicating that these biases were not statistically significant.

Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

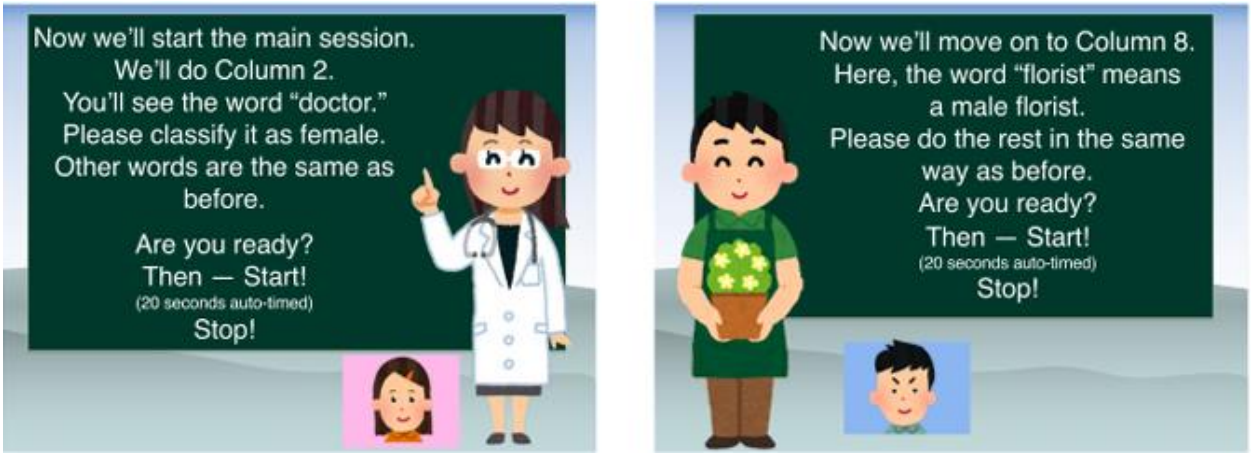


Figure 1.
Example slides illustrating a "female doctor" (left) and a "male florist" (right).

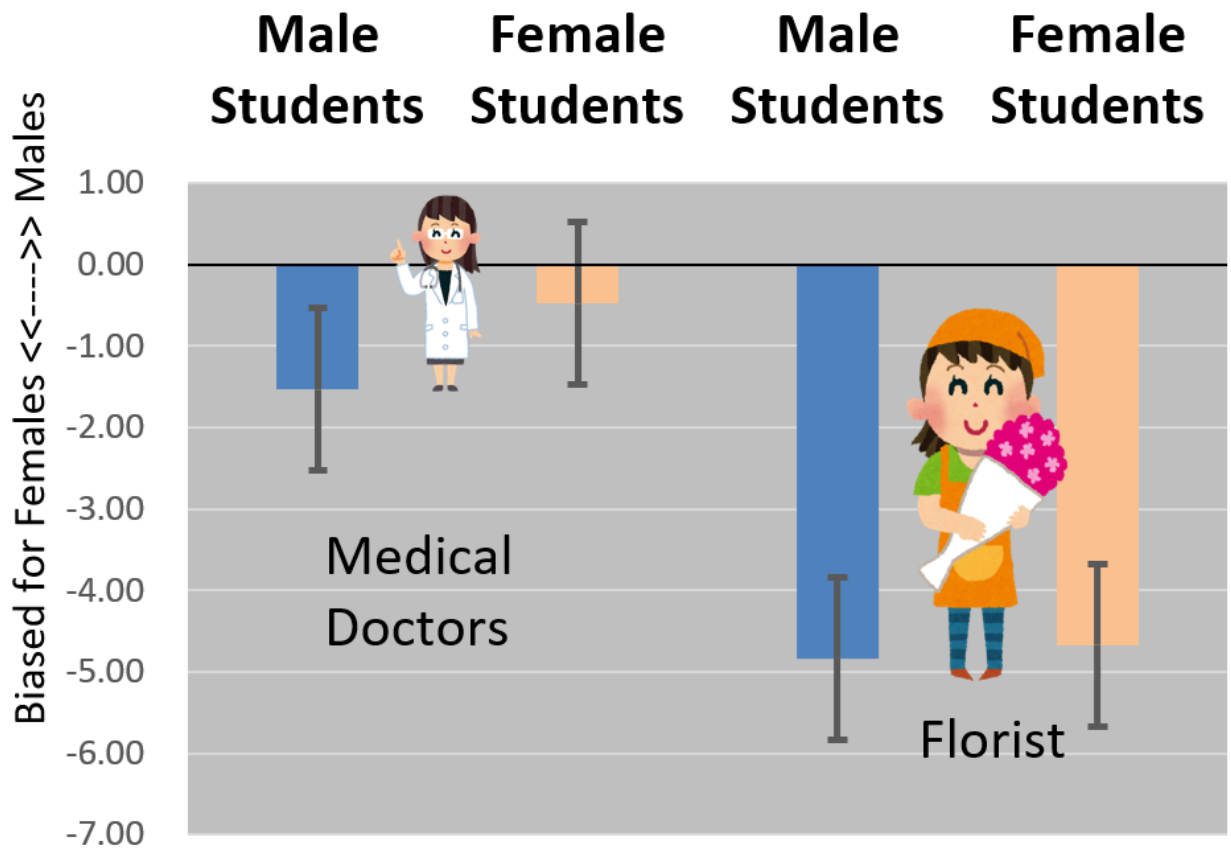


Figure 2.
Means and the 95% Reliability Ranges of the IGB100 Scores of Male and Female Students for "Doctor" and "Florist." (The vertical bars show the 95% reliability range.)

In contrast, for "florist", both male and female students showed notably negative IGB_{100} values (male students: −4.84; female students: -4.67). The upper limits of their 95% confidence intervals were also below

Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

zero ($-6.73 <$ male $IGB_{100} < -2.94$; $-6.09 <$ female $IGB_{100} < -3.25$), indicating that the gender stereotype "florist = female" was statistically significant.

6. Discussion

The gFUMIE test developed in this study successfully detected the gender stereotype "Florist = Female" among both male and female students. Our pilot study, Mori et al. (2025), also found this stereotype when administered to 44 college students majoring in nursery education. In contrast, the expected stereotype "Doctor = Male" was not detected. Mori et al. (2025) also failed to identify this doctor-related gender stereotype. These findings suggest that such stereotypes may no longer exist among contemporary young generations in Japan.

Why does the stereotype "Florist = Female" persist while "Doctor = Male" seems to have faded? In Japan, the number and proportion of female doctors have steadily increased. In 2022, female physicians exceeded 80,000, with their share rising to 23.6% (Nippon.com, 2024). Despite this, most doctors remain male. Conversely, florists are classified under retail trade in official statistics, and specific gender ratios are unavailable. However, private sources indicate women constitute a clear majority, with ratios from 1:9 to 25:75 in favor of females (Careergarden, 2025; Nicolai Bergmann Flowers & Design, 2025). Given these similar circumstances, why has the doctor stereotype disappeared while the florist stereotype has persisted?

One possible hypothesis is that gender equality education promoted by Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) has explicitly aimed to eliminate occupational gender stereotypes (Akita & Mori, 2025). Some occupations, such as florist, nurse, and nursery teacher, have traditionally been regarded as implicitly female. However, the primary targets of such educational efforts have been professions traditionally dominated by men, such as doctors and politicians.

For occupations requiring professional qualifications, such as nurses and nursery teachers, legal revisions changed their job titles from female-specific forms (kangofu and hobo) to gender-neutral ones (kangoshi and hoikushi). However, no similar reforms or stereotype-reduction measures have been applied to florists. To test this hypothesis, current occupational–gender stereotypes for professions like nurses and nursery teachers should be examined using gFUMIE, as in the present study.

Future studies should administer the gFUMIE test to participants from a broader range of age groups to examine age-related differences in the gender stereotype of doctors. Such investigations may reveal that the stereotype associating doctors with men has gradually diminished over time.

The children's version of the gFUMIE test may be used in future studies to examine how stereotypes develop across age groups. To do this, it is necessary to administer the test to lower-grade elementary school children and verify its validity. Once the effectiveness of the children's version is confirmed, the gFUMIE test could be applied to a broader range of professions beyond doctors and florists. This would enable researchers to investigate generational and developmental changes in occupational gender stereotypes.

7. Conclusions

To investigate the development of gender stereotypes, it is necessary to develop a measurement method that can be administered even to lower-grade elementary school children. In this study, we modified the FUMIE test and constructed a children's version of the gFUMIE test. Before verifying its validity with children, we conducted a preliminary validation by administering the adult version of the gFUMIE test to university students to examine whether gender stereotypes could be detected for the occupations of doctor and florist.

The results from 70 university students indicated that the gFUMIE test successfully detected a feminine gender stereotype for "florist." In contrast, the absence of the "Doctor = Male" stereotype among the same participants suggests this stereotype may have diminished among the younger generation. Future studies should administer the children's version of the gFUMIE test to lower-grade students to confirm its validity and explore developmental aspects of gender stereotypes in early childhood.

Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study, with no vital features omitted. Any discrepancies from the planned study have been explained. This study adhered to all ethical practices during writing.

Copyright:

References

Akita, S., & Mori, K. (2022). How implicit image of woman changed in Japanese sixth-grade children after a gender equality education lesson. *The Journal of Social Studies Research*, *46*(2), 153-159. https://doi.org/10.1016/j.jssr.2021.05.004

Akita, S., & Mori, K. (2023). Implicit image of women improved in Japanese children after gender equality education and maintained three years later. *SN Social Sciences*, *3*, 69. https://doi.org/10.1007/s43545-023-00655-1

Akita, S., & Mori, K. (2025). An experimental assessment of gender equality lessons for Japanese 4th graders. *Research in Social Sciences*, *8*(3), 1-10.

Akita, S., Tsushima, S., Saito, T., & Mori, K. (2019). The composition and administration of a group performance implicit association test for sixth grade elementary school students. *Research Bulletin of Matsumoto University*, *17*, 165-173.

Careergarden. (2025). *How to become a florist. Careergarden.jp*. Retrieved from https://careergarden.jp/hanaya/naruniha/

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology*, *109*(2), 187-200. https://doi.org/10.1016/j.jecp.2010.11.002

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Mori, H., Akita, S., & Mori, K. (2025). Development and preliminary validation of the gFUMIE test for measuring implicit gender stereotypes in young children: A pilot study with nursery education college students using "doctor" and "florist". *Bulletin of Bunka Gakuen Nagano Technical College*, *17*, 3-12.

Mori, K., Matsumoto, E., Karasawa, I., Ejima, M., Mure, A., & Uchida, A. (2019). Composition of a group performance implicit association test for third-grade elementary school children. *Research and Studies in Education, Matsumoto University*, *3*, 165-174.

Mori, K., Uchida, A., & Imada, R. (2008). A paper-format group performance test for measuring the implicit association of target concepts. *Behavior Research Methods*, *40*, 546-555. https://doi.org/10.3758/BRM.40.2.546

Nicolai Bergmann Flowers & Design. (2025). *Experience the Excitement of creativity — join the world of Nicolai Bergmann! Nicolai Bergmann Careers Page*. Retrieved from https://xl0yzvx08.jbplt.jp

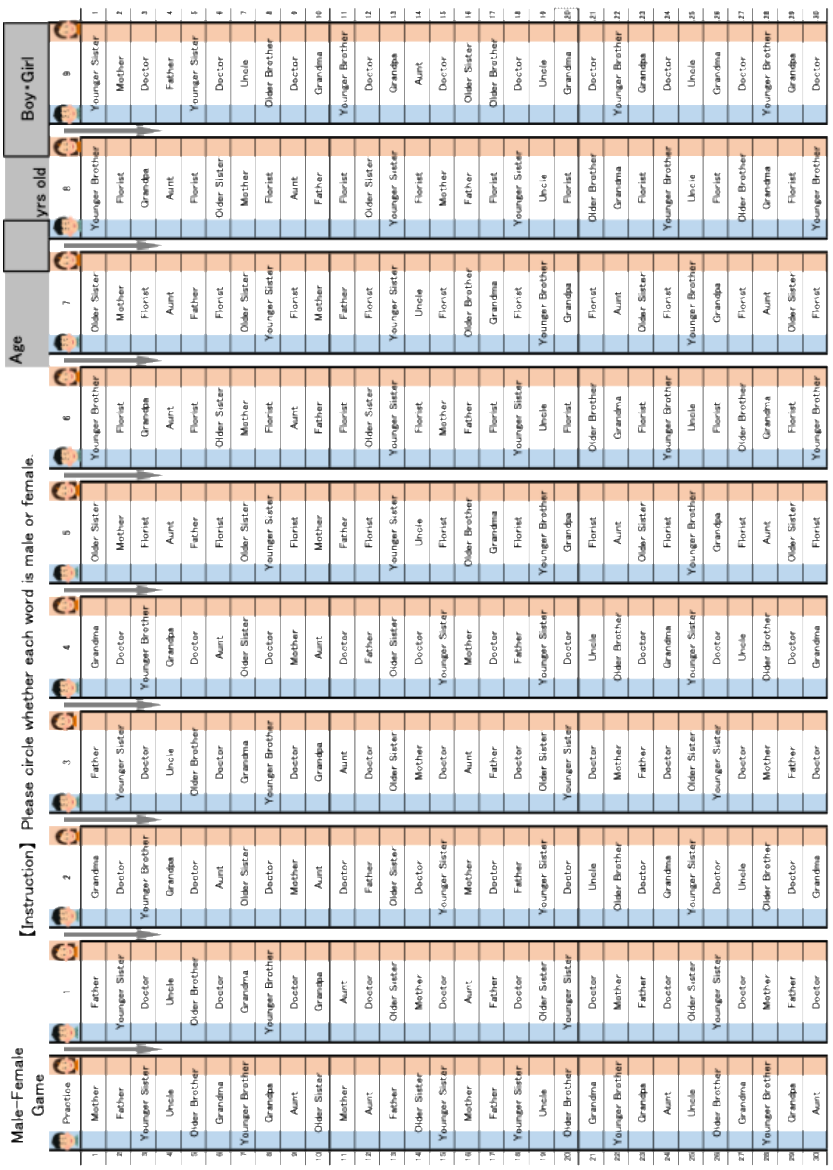
Nippon.com. (2024). *Record 80,000 female doctors in Japan, but huge gender gap remains, Japan Data, Nippon Communications Foundation*. Retrieved from https://www.nippon.com/en/japan-data/h01978/

Thomas, S., Burton Smith, R., & Ball, P. (2007). Implicit attitudes in very young children: An adaptation of the IAT. *Current Research in Social Psychology*, *13*, 75–85.

Uchida, A., & Mori, K. (2018). *FUMIE test administration manual (ver.2.2). do japanese junior high school students really dislike mathematics?: For promotion of the evidence-based education*. Kyoto, Japan: Kitaooji Shobo Publishing.

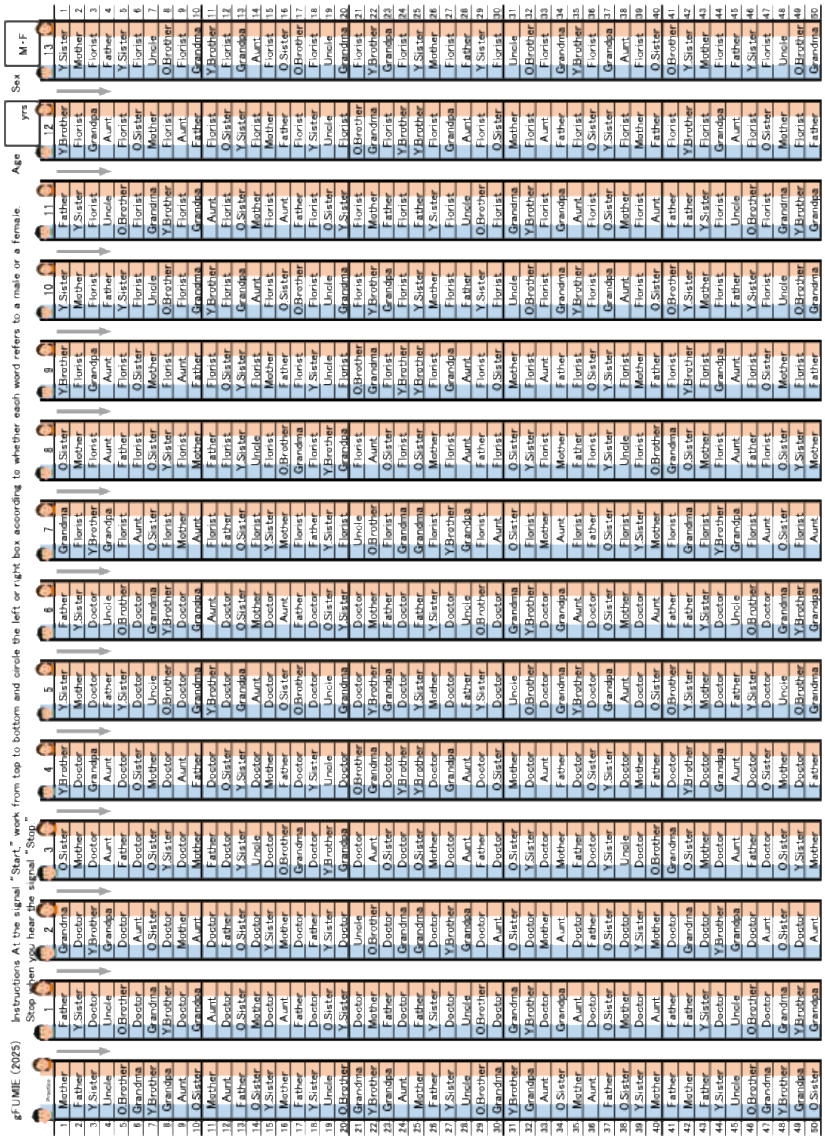
Whitford, A. (2023). *Investigating the implicit: An exploration of elementary students' conceptions of gender. In S. L. DeZutter (Ed.), International Perspectives on Educating for Democracy in Early Childhood: Recognizing Young Children as Citizens*. New York: Routledge.

Appendix A. The Boy-Girl Game Test Sheet (gFUMIE for Younger Children)



Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818

Appendix B: gFUMIE Test Sheet (Adult Version)



Journal of Contemporary Research in Social Sciences
ISSN: 2641-0249
Vol. 8, No. 1: 18-26, 2026
DOI: 10.55214/2641-0249.v8i2.11818