E-Journal of Philosophy of Education:
International Yearbook of the Philosophy of Education Society of Japan, Vol. 8, 2023

The applicability of the 'ethnometric method' as a support tool for teachers' self-evaluation: Focusing on children's physical expressions related to 'personalization and empathizing' in moral education[1]

Nobuo FUJIKAWA
Osaka University

Abstract

The ethnometric method is a method of generating evidence to support educational practitioners' subjective – or rather, intersubjective – impressions regarding the success or failure of their teaching practice by quantifying their evaluation of students' physical expressions. A main consideration when developing this method was utilizing Evidence-Based Education (EBE) as a starting point without succumbing to technocratic tendencies, and while respecting practitioner's agency.

The ethnometric method uses interviews to ask the practitioner what changes in physical expression he or she tends to focus on when evaluating the degree to which teaching goals were achieved. The practitioner is also asked to set up a video camera close to his or her own outlook on the classroom to film the teaching practice. The researcher then converts the filmed video into still images at intervals of a few seconds and counts the physical expressions that the practitioner focuses on as indicators for evaluation. The chronological change of the numerical data acquired through this method allows practitioners to substantiate or amend their subjective – intersubjective – impressions regarding the success or failure of their teaching practice.

This paper presents an ethnometric analysis of a series of Moral Education lessons in one fourth-

[1] This paper is part of the research results of Grant-in-Aid for Scientific Research, Basic Research (C) (General), Issue No. 21K02170, "Dotoku Jugyo Bunseki e no Tekiyo o tsujita 'Ethnometry-ho' no Yuyosei to Datosei no Kensho" [Verification of the usefulness and validity of the 'ethnometric method' through application to moral class analysis] (FY2021 - FY2023; Principal investigator: Nobuo Fujikawa). This research has also been approved (21007 and 22006) by the Research Ethics Committee of Osaka University, Graduate School of Human Sciences, Education.

Correspondence:
Nobuo FUJIKAWA, Osaka University. Email: fujikawa@hus.osaka-u.ac.jp

grade primary school class on the topic 'personalization and empathizing'. Comparing the results of the analysis with practitioners' impressions regarding the success or failure of achieving 'personalization and empathizing', the study finds that the ethnometric analysis largely supports practitioners' self-evaluation of their own classes.

Of course, this approach still faces several challenges. One criticism concerns the danger of this method's procedures leading to increased control and regulation of the body. Another criticism concerns limitations as to the range of circulation and validity of the evidence produced because the evidence is limited to social settings where practical knowledge or 'habitus' can be shared intersubjectively. However, such limited validity can also be an advantage because it has the potential of avoiding the danger of increased control of practice.

Key words: moral education, physical expression, personalization and empathizing, self-evaluation, evidence-based education

Introduction

The 'ethnometric method' is a method of generating evidence to support educational (and welfare) practitioners' self-evaluation of practices and outcomes[2]. This method, currently under development by the author, uses video (analysed as a series of still images) to quantify and visualize changes in participants' (e.g., students) physical expressions (e.g., facial expressions and gestures) and in this way allows educational practitioners to substantiate or amend their subjective – or rather, intersubjective – impressions regarding the success or failure of their teaching practice.

A main consideration when developing this method was utilizing Evidence-Based Practice (EBP; of which Evidence-Based Education [EBE] is a part) as a starting point without succumbing to technocratic tendencies, and while respecting practitioners' agency and intersubjectivity. The importance of a balance between 'research evidence', 'clinical expertise' and 'patient values' has been illustrated in Evidence-Based Medicine (EBM), which provided a major impetus for the development of EBP (Figure 1[3]). Applying this to EBE would stress the importance of a balance between 'research evidence', 'teachers' expertise' and 'students' values'. However, with Exam-orientation (in Japanese also termed *Hensachi*, that is, deviation-orientation)

---

[2] The author has previously used this method for the evaluation of welfare practice, in this paper the same method is applied to the evaluation of educational practice.

[3] Prepared based on Sackett, D. L. et al. 2000, p. 1.

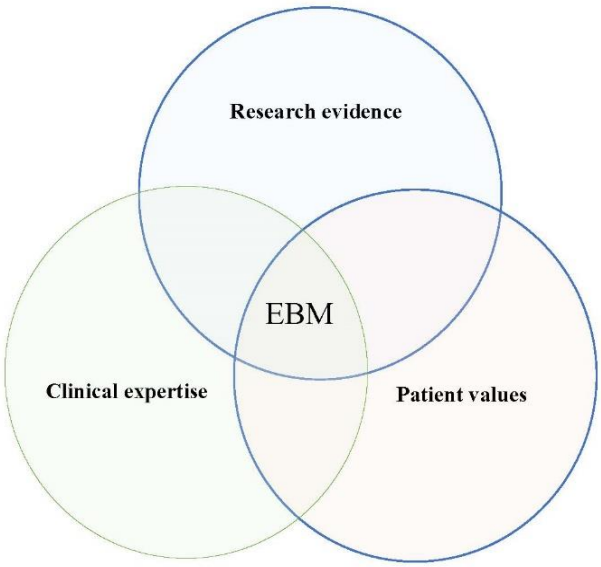


Figure 1: Components of EBM

permeating not only education but daily life, one might surmise that the latter two – 'teachers' expertise' and 'students' values' – which have already been worn thin by the orientation towards scientific evidence, will eventually lose their balancing function altogether. If this is the case, some measures are needed to more actively support the development of these latter two factors.

Teachers' judgment of the outcomes of their own teaching practices is based not only on the results of paper tests, but also on 'teachers' practical knowledge' (that is, cognitive and behavioural schemes or 'habitus' working at a semi-conscious level) [Gebauer, G. 2017]. This is an important component of teachers' expertise. 'Teachers' practical knowledge' is formed within daily interactions with fellow teachers and students over a long period of time. Thus, it is relatively stable, but continually and gradually changes, as it is also influenced by the expression of 'students' values'. Is it possible, then, for a person who is not a teacher to share in such 'practical knowledge' and to participate in examining its validity? If public trust in teachers were high enough, such would likely not be necessary. However, when considering the growing preference for evidence-based approaches, teachers' self-evaluations may be disregarded as mere 'subjective impressions', thus making an additional external perspective necessary. Interviews and questionnaires of teachers and students provide such perspective, however, as these research methods are highly dependent on language, they do not always fully able to encompass the impressions and intuitions that are an inherent part of 'teachers' practical knowledge'. For example, evaluation methods that are overly dependent on language will never be highly reliable, particularly when it comes to teachers' self-evaluation of subjects such as the Arts and Physical Education, but also in relation to self-evaluation of teaching in the lower primary grades in general[4].

[4] The large number of previous studies on physical expression in pre-school education, particularly in the areas of childcare and physical education, where it is considered difficult to evaluate lessons relying on language media, may be related to this. For reference, a CiNii search using the keywords '*Shintai Hyogen* (physical expression)' and '*Kyoiku* (education)' identified 550 papers, 33 books and 7 doctoral dissertations; for '*Shintai Hyogen*' and '*Hoiku*

Taking this into consideration, the 'ethnometric method' focuses on physical expression such as gestures and facial expressions. Bodily movements acquire meaning and function as physical expression through long-term participation in daily interactions within a particular social setting. It seems safe to assume that teachers, in evaluating their own teaching, base these evaluations not only on information derived from the linguistic activities of students, but also on information derived from their body language and expressions. The ethnometric method introduced here measures the frequency of the appearance/observation [5] of such physical expression and visualizes its transition in numerical and graphical form.

In the ethnometric method, the practitioner first sets up a (fixed) video camera close to his or her own outlook on the classroom and films the teaching practice (lessons, workshops, etc.). Next, in a short interview taking place after several lessons, the practitioner is asked to present participants' (in this case, students') physical expression (indicators) that he or she tends to focus on when evaluating the degree to which teaching goals were achieved. This would sum up practitioners' role in the implementation of the ethnometric method. Indeed, practitioners' relatively lighter burden at the data collection stage is also an important characteristic of this research method.

The researcher then converts the filmed video into still images at intervals of a few seconds. Focusing on a few minutes before and after interactions indicated in the practitioner's presentation, the researcher counts the frequency of appearance/observation of indicators of participants' physical expression. The figures derived from this analysis, and changes exhibited over time, provide practitioners with a means to visualize the intersubjective impressions they have concerning the outcome of particular teaching experiences, as well as lengthier series of practices. This is the main procedure entailed by the ethnometric method.

(childcare)', 406 papers, 18 books and 3 doctoral dissertations; for '*Shintai Hyogen*' and '*Taiiku* (physical education)', 284 papers and 10 books; for '*Shintai Hyogen*' and '*Ongaku Kyoiku* (music education)', 74 papers and 6 books; '*Taiiku*' and '*Yoji Kyoiku* (early childhood education)' yielded 64 papers and 15 books; for '*Shintai Hyogen*' and '*Bijutsu Kyoiku* (art education)', 4 papers (search date: 5 October 2022).

[5] It is important to note that the while this method allows a researcher, by partially sharing the perspective of the classroom practitioner, access to students' physical expression, only the physical indicators that are recognizable through the medium of the still image can be recorded. Thus, the output derived from these measurements by no means fully reflects students' realities (in fact, such cannot be grasped by anyone, not even by a highly developed AI). Rather, this method records only some of the 'intersubjective' meanings accessible within the interactive environment of the classroom. To indicate these limitations, this paper utilizes the term 'frequency of appearance/observation'.

Of course, this approach still faces several challenges. Firstly, a fundamental criticism concerns the danger of this method's procedures leading to increased control and regulation of the body. In this sense, the ethnometric approach is also not free from what Campbell termed 'corruption pressure'[6] in regards to evidence[7]. Secondly, there is also the issue of costs in terms of the time, people and money required to implement this method in educational settings. Due to the required resources, the method can likely only be used in settings such as a funded field-research project, therefore at present, a rubric evaluation[8] remains a more feasible research method. The solution is to be expected from the future development and dissemination of AI[9]. There is, however, a danger that such expectations will invite further 'corruption pressure'.

Nevertheless, according to the ethnometric method, the indicators to be measured (participants' physical expressions) are set based on practitioners' intersubjective impressions. Therefore, the range of circulation and validity of the evidence produced is limited to a social setting where practical knowledge or 'habitus' can be shared intersubjectively (in the case of education, the scale of in-school training for instance). If this is the case, then the perceived disadvantages in terms of a limited scope and the intersubjectivity of the evidence may in fact become an advantage in terms of avoiding 'corruption pressure' and the danger of increased control of practice. The word 'ethno' in ethnometric implies exactly such intentions, or rather aspirations.

The author has previously implemented the ethnometric method to evaluate a theatre workshop at an elderly care facility and a contemporary dance workshop for young children at a community centre [Fujikawa, N. 2020]. This paper reports on the first time the method has been used to evaluate teaching in schools. The following section will present (1) an overview of the research subjects and the concrete measurements and analysis procedures conducted, followed by

---

[6] According to D. T. Campbell, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" [Campbell, D. T. 1975, p. 35].

[7] Similar to this paper, Lind's paper also takes the position that EBE is useful, with some limitations, for self-evaluation by practitioners. In particular, it considers 'corruption pressure' in cases in which 'research evidence' is utilized in relation to educational administration [Lind, G. 2011].

[8] While it is possible for a rubric evaluation to emphasize and score the intersubjective impressions of outside observers, practitioners, and also participants, the ethnometric method is unique in that it specializes in physical expression and allows to count the frequency of its appearance/observation more precisely based on still images.

[9] As research in the field of educational technology suggests, AI is likely to take on the task of data analysis in the future [Shibata, Y. 2021].

(2 and 3) partial results, and (4) a discussion of the efficacy of the ethnometric method. Lastly, (5) the author will deploy a Science Fiction of Education (SFE) to imagine the future of 'measurement' and discuss the meanings encompassed by the term 'ethno'.

1. Aims, subjects and procedures

(1) Research goals (Research questions)

This research aims to answer the following three questions.

(i) *Did the frequency of appearance/observation of students' physical expressions (indicators) increase significantly after a teacher puts forth questions designed to encourage 'personalization and empathizing' (in Japanese: 'Jibungotoka')*[10] *in Moral Education lessons?*

(ii) *Has veteran teachers' teaching practice functioned as a model for younger teachers?*

(iii) *Can the results produced by applying the ethnometric method's analysis support teachers' self-evaluation?*

(2) Research subjects

The subject of this research was a series of Moral Education lessons held from 22 April to 15 July 2021 in one fourth-grade class at a public primary school in Shimane Prefecture, Japan (45 minutes per lesson)[11]. In this series of lessons, teacher A, a younger female teacher, was responsible for the first to fourth and seventh to eighth lessons, while teacher B, a veteran male teacher, was responsible for the fifth and sixth lessons (In lessons taught by teacher A, teacher B was observing; and vice versa, in lessons taught by teacher B, teacher A was observing). The fourth session was excluded from the research due to a malfunction of the filming equipment. Therefore, the data for measurement and analysis was comprised of a total of eight lessons. This paper presents interim results from this study, based on the data from the 3rd, 5th, 6th and 7th lessons, which have been analysed at this stage. The specific subject of measurement was

---

[10] In the context of moral lessons, '*Jibungotoka*' refers to the ability to empathize with characters in stories that serve as teaching materials, understanding the events that happened to them as happening to oneself, and thinking about and making judgements about those events through one's own experiences and with a sense of ownership. In this paper, the terms 'personalization and empathising' will be used as a translation of '*Jibungotoka*'.

[11] Although their interests and methods differ from those of this paper, for a summary of the results of these series of lessons refer to Otsuka, T. & Shiotsu, H. 2022.

students' physical expression during the five minutes before and after the teacher put forth a discussion question[12] designed to encourage 'personalization and empathizing' in each lesson. Table 1 presents the following details: date and starting time of each lesson, teacher responsible, textbook teaching unit, and question to encourage 'personalization and empathizing' for each lesson. In these lessons, 'personalization and empathizing' is defined by veteran teacher B as 'imagining the living environment surrounding oneself and reconstructing one's values in order to reflect on one's way of life and behaviour, while studying the contents of each learning unit' [Otsuka, T. & Shiotsu, H. 2022, p. 28].

[12] The questions meant to encourage further 'personalization and empathizing' in these lessons are positioned as supplementary questions posed after the central question of each teaching unit [Otsuka, T. & Shiotsu, H. 2022, p. 28].

Table 1: Each lesson's responsible teacher, teaching unit, and 'personalization and empathizing' questions

Lessons	**Date (starting time)**	**Responsible teacher**	**Teaching unit from *"Dotoku 4: Kimi ga ichiban hikaru toki"* (*Moral 4: When you shine the brightest*) (Mitsumura Book Co., Ltd.)**	**'Personalization and empathizing' questions (time)**
1	April 22. (10:55-)	A	"Saturday School"	Why is it important to be polite? (11:25)
2	May 6. (10:55-)	A	"What is 'consideration'?"	What is 'consideration'? (11:26)
3	May 10. (10:51-)	A	"Picture postcards and stamps"	What does it mean to "cherish your friends"? (11:22)
5	June 3. (10:57-)	B	"Alarm Clock"	How should I continue to do 'what I can by myself'? What feelings are important? (11:25)
6	June 10. (10:55-)	B	"A Sign of Life"	When you were born, how did the people around you (your family members) feel? (11:31)
7	June 17. (10:57-)	A	"The Lonely Student Y-chan"	Why is it important to treat everyone the same? (11:23)
8	July 1. (10:56-)	A	"If we had just left it like this"	Do you ever feel like this? ...Do you ever feel like 'I' do, thinking "Oh, what a hassle!"? (11:31)
9	July 15. (10:57-)	A	"We'll be waiting for you"	Have you ever been encouraged by your close friends and felt "I'm feeling better now!" or "I'm going to do my best!"? (11:22)

(3) Procedures

The procedures for data collection, measurement and analysis were as follows:

(i) The teacher set up two video cameras close to her/his own position (at the front of the classroom, on either side of the teacher's desk; each camera's perspective was different but they partially overlapped) and started filming at the start of the lesson.

(ii) The researcher (Fujikawa) conducted a short interview by email after the lesson series had been completed [August 6, 2021]. The two teachers discussed indicators of students' physical expression that they both focused on when evaluating the achievement of 'personalization and empathizing'[13]. As a result, they defined seven indicators as listed below (see Table 4 appended at the end of this paper for indicators' selection rationale).

Table 2: Indicators of physical expressions for measurement and analysis

	Concrete expressions
Indicator 1	Nodding (to teacher's statements)
Indicator 2	Leaning forward (excluding cases involving raising hands)
Indicator 3	Widening eyes (lively gaze)
Indicator 4	Upward gaze (a short pause in response to a question and staring into space)
Indicator 5	Fixed gaze (gaze stops moving in response to the question posed by the teacher)
Indicator 6	Conversation starts 'right after' encouraging discussion with neighbouring students in the course of the lesson
Indicator 7	Various neck movements such as nodding, shaking one's head from side to side and tilting the head sideways appear in succession

(iii) The researcher converted each of the five-minute video clips before and after the question that

[13] In this study, the two teachers who were responsible for the classes happened to discuss and decide together which indicators will be measured and analysed. However, as an after-the-fact reflection, I believe that this point is very important and should therefore have been incorporated into the research design from the outset. The importance of discussion would become clear if one considers what would happen if a teacher were to cite 'children sleeping soundly' as an indicator for a great lesson. That is, we should not assume that all teachers believe that human nature is fundamentally good. Thus, it may be desirable to involve one or more people (ones who are invested in students' care and success) alongside the practitioner in setting the indicators. This point is also related to the discussion on "deliberation" at the end of this paper.

prompted 'personalization and empathizing' into still images at six-second intervals (approximately 50 images of each 5 minutes before and after the question was asked). The still images were used to count the frequency of appearance/observation of the students' physical expressions (indicators). In order to account for measurement errors, the images taken by one camera were counted by a total of four persons acting as measurers (a pair of university students, their resulting count was termed 'normal accuracy' data, and another pair who were people considered to be particularly sensitive to physical expression such as theatre actors, their output was termed 'high accuracy' data)[14]. In order to get as close as possible to the teachers' subjective impression of the lesson, two counts were conducted. In the 'unweighted' count, the appearance/observation of an indicator in any of the students was counted as 1 point. In the 'weighted' count, indicators of specific students' physical expression counted as 2 points[15]. These were students that were of particular concern to the teachers. The reasons for being 'of concern' were: "They fill in the 'reflection sheets' in short sentences, and do not usually say or respond much" (2 students); "They have some difficulties with writing, and therefore filling in the 'reflection sheets' is also challenging, however based on their responses in class we often see moral growth" (2 students); "The student's desire to learn is irregular and he/she rarely attempts to participate in classes they are not interested in" (1 student) [discussed and answered by both teachers in response to questions by email on 24 January]. To enable the measurers to identify these specific students, the teachers' attached arrows to some of the still images captured by the two cameras at each lesson.

(iv) The researcher carried out the following analytical work based on the data obtained in (iii) above.

(a) The 'amount of physical expression per person' was calculated by dividing the points indicating the appearance of each physical expression (indicator) by the number of photographic subjects in each still image (approximately 50 photographs of each 5 minutes before and after the 'personalization and empathizing' question in each lesson were eligible for the count).

(b) Based on the data in (a), a line chart was created with the 'amount of expression per person' on the vertical axis and the 'still image serial number (= time)' on the horizontal

[14] Presumably the 'high accuracy' data is closer to the impressions and perception of skilled teachers.

[15] The 'weighted' count was distinguished from the 'unweighted' count in order to see if a significant difference is noticeable in graph form or as a numerical value. The value of the 'weighted' count was provisionally set as 2 for convenience, not for any specific reason.

axis.

(c) Based on the data in (a), a box-and-whisker diagram showing the highest value, 25% above the median, the median, 25% below the median and the minimum value was prepared.

(d) Based on the data in (a), a Brunner-Munzel test was conducted for the five minutes before and after the 'personalization and empathizing' question, with the null hypothesis of 'no difference between the two groups before and after'.

This concludes the basic steps of the ethnometric method. The results produced by the above procedures provide an answer to research question (i). Comparing the figures and graphs obtained from the lessons given by the younger teacher A before and after those given by the veteran teacher B will provide an answer to research question (ii). In addition, research question (iii) further aimed to explore whether the results of the analysis by the ethnometric method can support teachers' own impressions of lesson outcomes. To provide an answer the two teachers were also asked to give a self-evaluation on a 10-point scale in regards to the following question: "In each lesson, how well do you think the subject of 'personalization and empathizing' went? [conducted January 24, 2022] Table 3 details the results.

Table 3: Teachers' 10-point self-evaluation of each lesson

Lesson no.	Teacher responsible	Teachers' self-evaluation (10-point scale)
1	A	2
2	A	2
3	A	2
5	B	8
6	B	6
7	A	3
8	A	3
9	A	2

2. Measurement results

Figure 2: 3rd lesson, camera 1, normal accuracy, unweighted, 5 min. before the Q

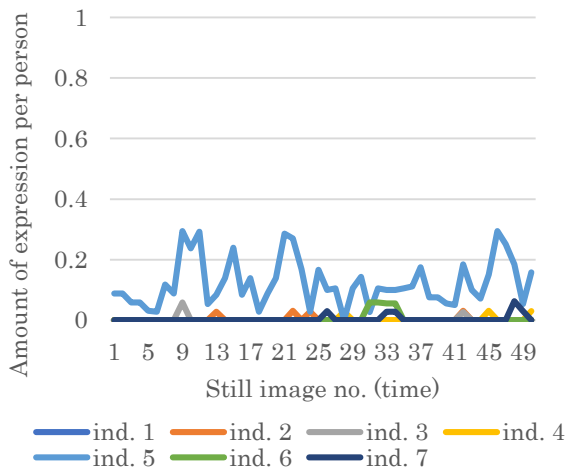
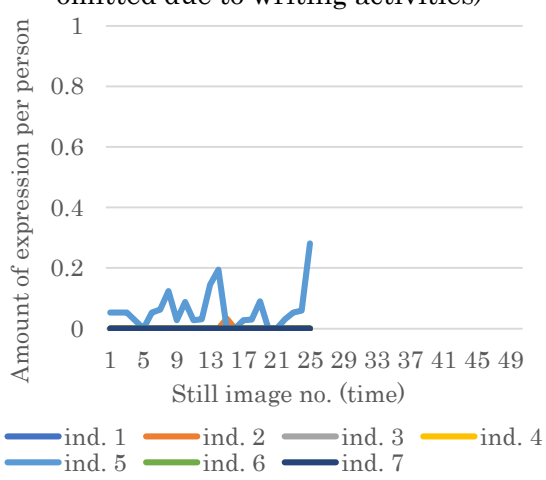


Figure 3: 3rd lesson, camera 1, normal accuracy, unweighted, 5 min. after the Q (photos 26-50 omitted due to writing activities)



First, based on the data obtained in the third lesson, let us show the change in physical expression during the 5 minutes before and after posing the question to encourage 'personalization and empathizing'. Figures 2 and 3 detail the changes in the seven indicators in the case of 'camera 1, normal accuracy, unweighted'.

These figures show that the value of indicator 5 ('fixed gaze') is particularly high compared to other indicators. This shows that the frequency of appearance/observation of this indicator is high. Furthermore, the amount of physical expression for indicator 5 slightly decreases from five minutes before to five minutes after the question.

As Figures 4 and 5 show, when focusing on students that teachers were concerned with (camera 1, normal accuracy, *weighted*), the top of the line (vertical axis value) is slightly higher, but the shape of the line itself does not change significantly.

Figure 4: 3rd lesson, camera 1, normal accuracy, *weighted*, 5 min. before the Q

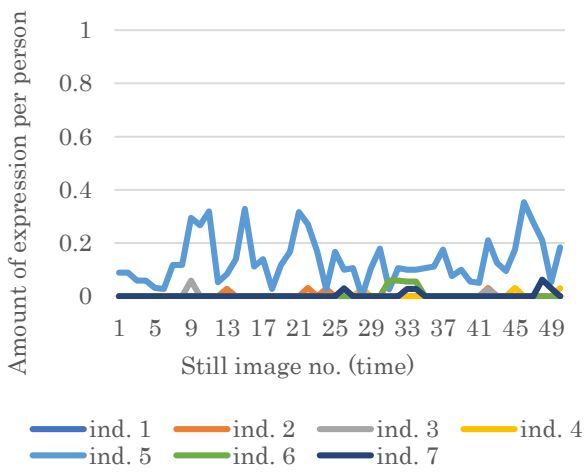
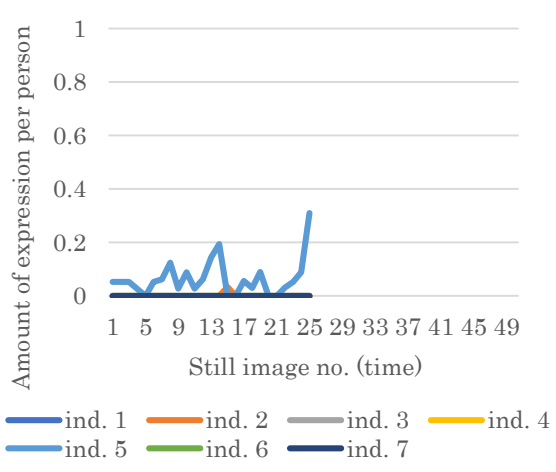


Figure 5: 3rd lesson, camera 1, normal accuracy, *weighted*, 5 min. after the Q (photos 26-50 omitted due to writing activities)



Figures 6 and 7 reveal that the data obtained from camera 2 (*camera 2*, normal accuracy, unweighted) present a similar trend as the data from 'camera 1, normal accuracy, unweighted' (Figures 2 and 3), even though the photographic subjects are different.

Table 6: 3rd lesson, *camera 2*, normal accuracy, unweighted, 5 min. before the Q

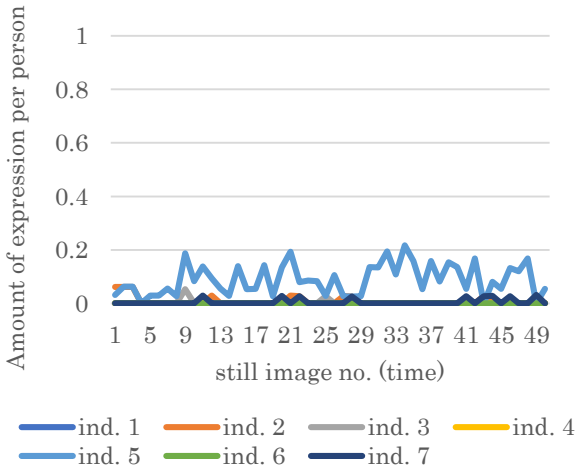
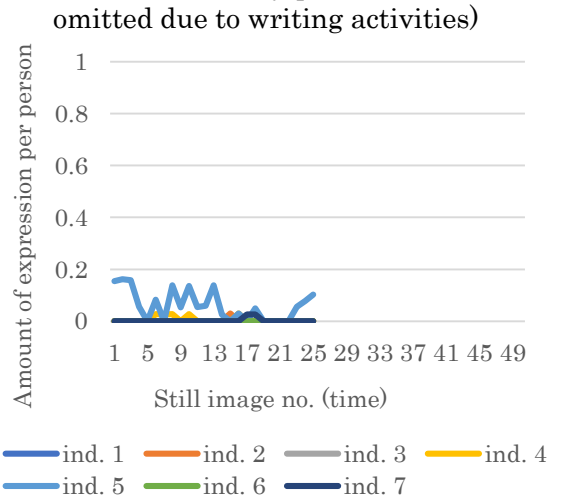


Table 7: 3rd lesson, *camera 2*, normal accuracy, unweighted, 5 min. after the Q (photos 26-50 omitted due to writing activities)



Furthermore, Figures 8 and 9 demonstrate the data of "camera 1, *high accuracy*, unweighted", where indicators were counted by a person considered to be sensitive to physical expression. Compared to the data of "camera 1, normal accuracy, unweighted" (Figures 2 and 3),

the frequency of indicator 5, as well as all other indicators, is slightly higher, and the difference in the height (vertical axis value) of the lines is also more apparent. This would suggest that more precise measurements (more sharply defined) were conducted. However, there was no difference in the trends (underlined) described above.

Figure 8: 3rd lesson, camera 1, *high accuracy*, weighted, 5 min. before the Q

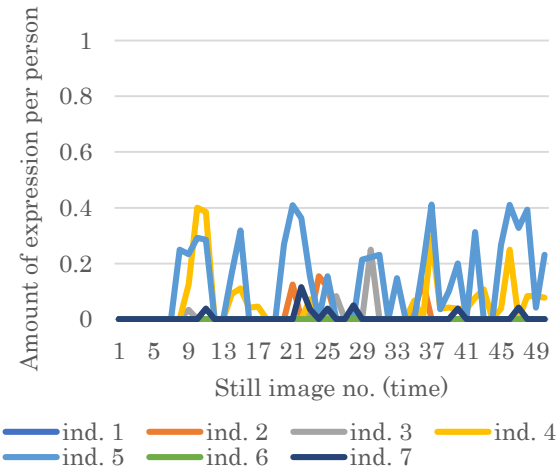
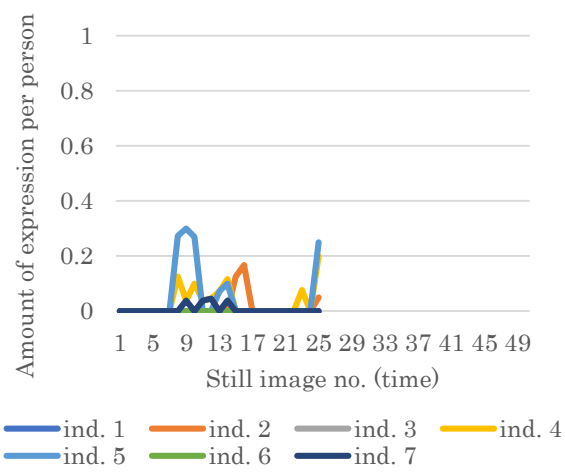


Figure 9: 3rd lesson, camera 1, *hight accuracy*, weighted, 5 min. after the Q (photos 26-50 omitted due to writing activities)



Furthermore, let us look at the changes in the amount of physical expressions for the *fifth*, *sixth* and *seventh* lessons, based on the data from 'camera 1, normal accuracy, unweighted'.

The line graphs for the *fifth* lesson taught by veteran teacher B (Figures 10 and 11) show that compared to the third and seventh lessons (Figures 2, 3, 14, and 15), the values of indicators 2

Figure 10: *5th* lesson, camera1, normal accuracy, unweighted, 5 min. before the Q

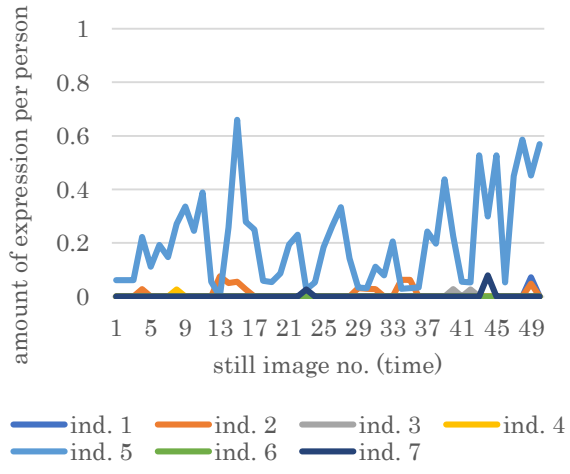
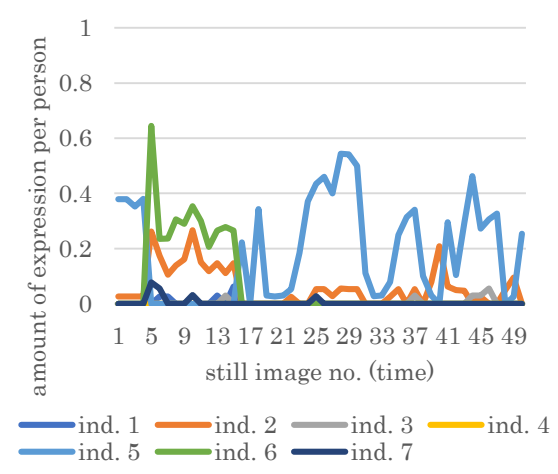


Figure 11: *5th* lesson, camera 1, normal accuracy, unweighted, 5. min. after the Q



('leaning forward'), 5 ('fixed gaze') and 6 ('conversation starts') are particularly high, both before and after the question. In addition, it can be seen that the values of indicators 2 and 6 clearly increased after posing the question to encourage the students' 'personalizing and empathizing', while the value of indicator 5 cannot be said to increase or decrease either way.

On the other hand, the line charts of the *sixth* lesson (Figures 12 and 13), which was again taught by teacher B, reveal that regardless of the timing before or after the question, the values of indicators 2 and 5 are higher than in the other lessons. Oddly enough, however, the values of both indicators decrease after posing the question. This point will be explained later.

Figure 12: *6th* lesson, camera 1, normal accuracy, unweighted, 5 min. before the Q

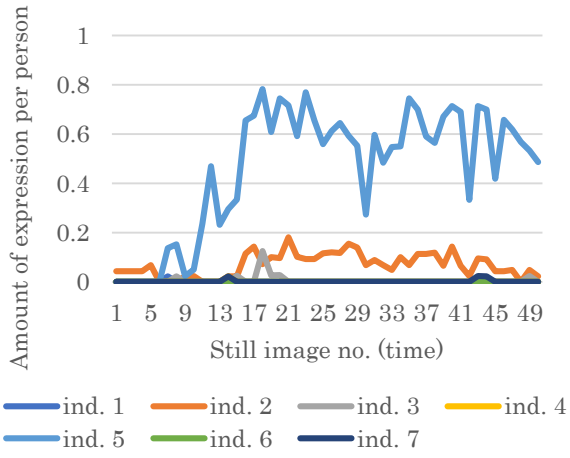
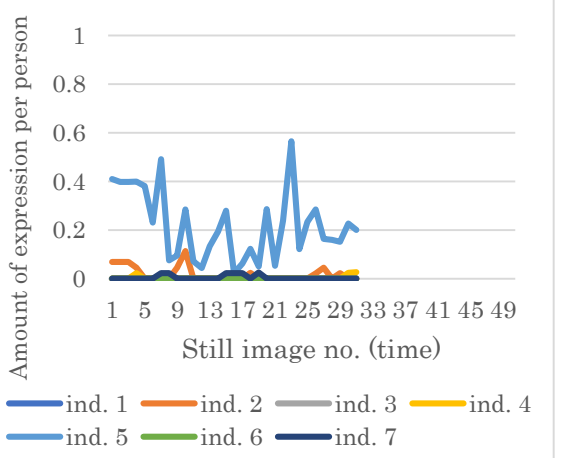


Figure 13: *6th* lesson, Camera 1, normal accuracy, unweighted, 5 min. after the Q (photos 32-50 omitted due to writing activities)



The graphs for the *seventh* lesson, which was taught by teacher A (Figures 14 and 15), show an overall decrease in the values of indicators compared to teacher B's lessons. Indicator 5 however does demonstrate an increase, especially after the question, when compared to the third lesson, which was taught by teacher A as well (Figures 2 and 3).

Changing the presentation method, let us now present a box-and-whisker diagram (Figure 16). This figure fucuses on the changes in values of *indicator 5*, which showed relatively high values in all the lessons, before and after the question ('camera 1, normal accuracy, unweighted'). The figure shows the numerical value of each indicator (amount of physical expression per person) on the vertical axis, and the differences between lessons, before and after the question, on the horizontal axis. The numerical rank order column was then divided into four parts; the top whisker represents the top 25%; the top of the box represents the 25% above the median; the horizontal line in the middle of the box represents the median; the bottom of the box represents the 25% below the median; and the bottom whisker represents the 25% below that.

Figure 14: *7th* lesson, camera 1, normal accuracy, unweighted, 5 min. before the Q

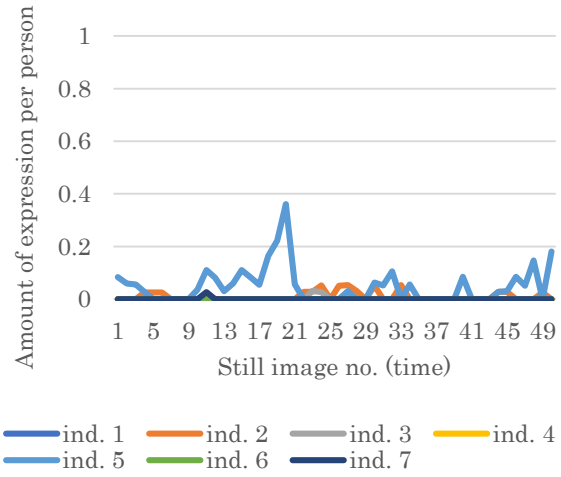


Figure 15: *7th* lesson, camera 1, normal accuracy, unweighted , 5 min. after the Q (photos 28-50 omitted due to writing activities)

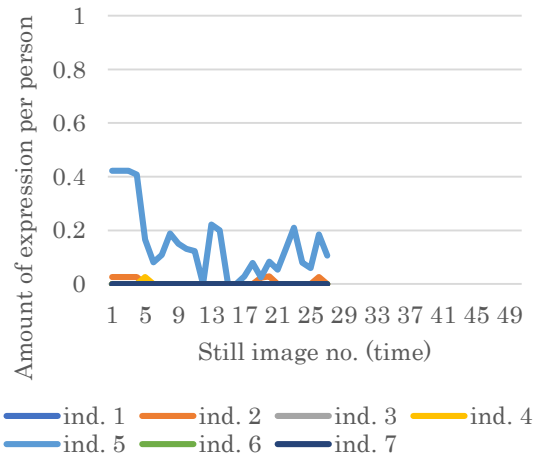
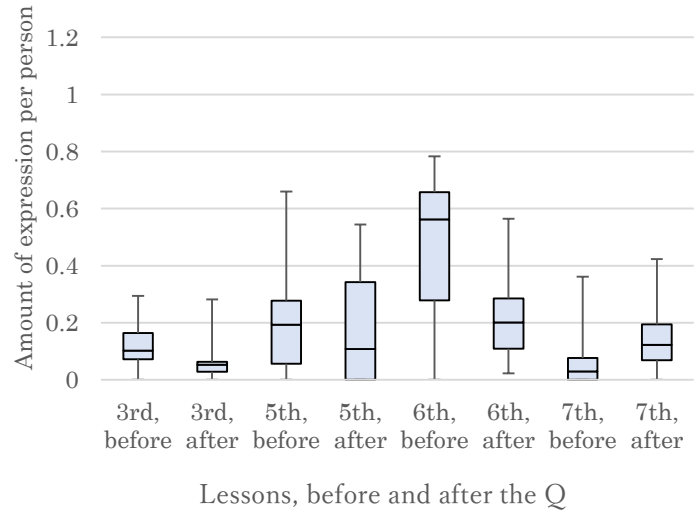


Figure 16: Data distribution of median and 25% above and below the median for indicator 5 for the 3rd, 5th, 6th, and 7th lessons, before and after the Q (camera 1, normal accuracy, unweighted)



This figure allows us to identify the finer points that could not be ascertained based on the line charts. First, a comparison of the *third* and *seventh* lessons shows a slight increase in the overall value of indicator 5 in the *seventh* lesson, both before and after the question. Moreover, whereas in the *third* lesson the values decrease after the question, they increase in the *seventh* lesson. In the *fifth* lesson by teacher B, the upper limit of 25% above the median value increases, whereas the median value decreases. Nevertheless, the values of indicator 5 in the *fifth* lesson were higher overall (both before and after the question) than those counted during the *third* and *seventh* lessons, which were taught by teacher A. In the *sixth* lesson, taught by teacher B, indicator 5 values were also higher overall (both before and after the question) than in the *third* and *seventh* lessons taught by teacher A, however there was a significant decrease in the highest and median values after the question.

3. Analysis results

Although this is only a small part of the data, the above figures allow an intuitive understanding of the approximate changes (especially in indicator 5) over the four lessons[16]. In the following, we will present the results of the Brunner-Munzel test. As previously described, the null hypothesis assumed there is no difference between the values measured during the 5 minutes before the question, and those measured after it. The test includes the measurement results of all indicators during the *third*, *fifth*, *sixth* and *seventh* lessons. The test results will be briefly discussed below based on the tables at the end of this document (Tables 5 and 6), which summarise the relevant data. In the tables, significant results from the tests are indicated by yellow and orange highlights. Based on the change in mean rank between 5 minutes before and 5 minutes after the question, results highlighted in orange indicate that frequency of physical expression indicators increased significantly after posing the question. Conversely, results highlighted in yellow indicate a significant decrease. In these tables, [**] marks instances of $p < .010$ and [*] marks instances of $p < .050$.

Let us first look at the results of the analysis in the case of '*normal accuracy*' data (Table 5). In the *third* lesson taught by teacher A, indicator 4 of camera 2 shows significant increases ($p = .035$* and $p = .014$*) before and after the question in both unweighted and weighted cases. However, other indicators in both camera 1 and 2 show no significant change, and some even show a significant decrease.

In the *fifth* lesson taught by teacher B, the results of the unweighted and weighted analyses are consistent for each camera, with significant increases in indicators 2 and 6 for camera 1 and only indicator 6 for camera 2 (both $p < .010$**).

In the *sixth* lesson by teacher B, the unweighted and weighted results are also almost identical for each camera. There is no significant overall increase in either camera, but rather a significant decrease in indicators 2, 3 and 5 for camera 1, and indicators 2 and 5 for camera 2 (both $p < .010$**).

In the *seventh* lesson taught by teacher A, data from camera 1 shows a significant increase in indicator 5 for both weighted and unweighted counting (both $p < .01$**). The data from camera 2 shows a significant increase in indicator 7 for the unweighted counting ($p < .010$**) and indicators 3 ($p < .010$**), 5 ($p = .046$*) and 7 ($p = .041$) for the weighted counting. It is worth noting that only in the weighted count of the data from camera 2 indicators 3 and 5 also showed significant increases.

---

[16] For teachers to use this data as a reference point for improving their teaching, it is probably not necessary to go as far as the statistical testing process, charting alone would likely suffice.

Let us now look at the results of the analysis for the '*high accuracy*' data (Table 6).

In the *third* lesson, indicator 4 shows a significant increase ($p < .010$**) in both weighted and unweighted counting for camera 2, while indicators 3 and 5 show significant decreases ($p = .039$* and $p < .010$** respectively) for camera 1.

In the *fifth* lesson, indicator 6 shows a significant increase, consistent across both cameras (both $p < .010$**).

There is no significant increase in the *sixth* lesson. Rather, indicators 2, 3, 4 and 5 show significant decreases for camera 1 and indicators 2 and 5 for camera 2 (both $p < .010$**).

In the *seventh* lesson, the values for both cameras do not differ in relation to on whether they were weighted or not. For camera 1, indicator 4 shows a significant increase ($p < .010$**), while indicator 3 shows a significant decrease ($p = .025$*). On the other hand, for camera 2, indicator 5 shows a significant increase ($p < .010$**) and indicator 4 a significant decrease ($p = .043$*).

4. Discussion

Before considering the three research questions presented at the beginning of this paper, it is first necessary to examine why the frequency of appearance/observation of some indicators significantly decreased in the *sixth* lesson conducted by teacher B in the five minutes after posing the question to encourage 'personalization and empathizing', although the teacher's self-evaluation was quite high (6 out of 10, see Table 3).

When the researcher asked teacher B about this question later [September 7, 2022], he said that in this lesson (teaching material 'A Sign of Life'), before posing the question to encourage 'personalization and empathizing', another prompt was given by the teacher. This activity required that "students who remember the birth of their younger brother or sister tell the class about their feelings at that time". Teacher B explained that, "This prompt probably enabled students to easily create a realistic image of the scene in their minds. Indeed, many students responded, 'I remember that!' or 'I don't have a younger brother or sister, so I don't know how I would feel. I'd like to hear how those who have a brother or sister feel!' Probably that's why many physical expressions appeared (before the question to encourage 'personalization and empathizing')".

Even when considering this irregular case, it can be said that the results of the ethnometric analysis can support teachers' self-evaluation of their own lessons (Table 3) by using data regarding the appearance/observation of students' physical expression.

Let us now consider the research questions posed at the beginning of this paper.

First, research question (i) *"Did the frequency of appearance/observation of students' physical expressions (indicators) increase significantly after a teacher puts forth questions*

designed to encourage 'personalization and empathizing' in Moral Education lessons?"

This question can be answered as follows. Analysis results demonstrated that the *third* lesson taught by teacher A was not necessarily successful in encouraging 'personalization and empathizing'; the *fifth* lesson taught by teacher B was quite successful in promoting 'personalization and empathizing'; during the *sixth* lesson taught by teacher B, 'personalization and empathizing' was in fact encouraged by another prompt before posing the question, therefore it cannot be accurately evaluated; in the case of the *seventh* lesson taught by teacher A, although it did not reach the level of the *fifth* lesson by teacher B, it can be said that 'personalization and empathizing' was promoted more than in the *third* lesson, which was also taught by teacher A; in addition, when comparing the unweighted and weighted data for the normal accuracy and camera 2, the weighted data shows a significant increase in indicators 3 and 5 after the question was put forth. This would suggest that the *seventh* lesson taught by teacher A may have contributed to the increase in students' physical expression without exception, that is, including the students that teachers were particularly concerned with.

Next, let us consider research question (ii): *"Has veteran teachers' teaching practice functioned as a model for younger teachers?"* Comparing the data collected during the *third* and *seventh* lessons taught by the younger teacher A, and the *fifth* and *sixth* lessons by the veteran teacher B, it is possible that the lessons taught by teacher B served as a model for the lessons by teacher A, which in turn reinforced teacher A's efforts to encourage 'personalization and empathizing'. The *sixth* lesson by teacher B may have also been a model for teacher A's lessons, including the supplementary prompt mentioned in teacher B's account of the class (that was posed before the question to encourage 'personalization and empathizing'). As the data showed, before the question several indicators were relatively high compared to the *third* lesson taught by teacher A.

Finally, research question (iii): *"Can the results produced by applying the ethnometric method's analysis support teachers' self-evaluation?"* Considering the results' correspondence with Table 3 (Teachers' 10-point self-evaluation of each lesson) this question can be answered in the affirmative[17].

---

[17] In this paper, the Brunner-Munzel test was conducted, but there is still room for improvement in the testing process, as there are disadvantages to transforming the data (number of physical expressions per person) into ranks and adding tests. To more accurately quantify data trends, it is probably preferable to apply Bayesian statistics. Bayesian statistics could, for example, divide the students into two groups, one group less likely to respond to the teacher's encouragement to 'personalization and empathizing' from the other group more likely to respond to it, and reveal how each group responded.

5. Where does a measurement-orientation lead and why do we want more precise measurements in the first place?

From the start, the design of the ethnometric method entails several constraints. First, it is the *practitioners themselves who define the physical expressions (indicators) to be measured*. This means that the indicators used in this study cannot be used for the measurement and analysis of other classes. The physical expressions defined as indicators by teachers can be understood as symbols that had accumulated their meaning through everyday interactions between students and teachers, and between colleagues. Naturally different indicators must be set for different teachers and classes. Second, the ethnometric method *does not seek to measure the reality of physical movements and internal states of participants.* The present research measured the movements that teachers, as a result of long-term everyday interactions with their fellow teachers and students, came to perceive as physical expressions signifying mental processes of 'personalization and empathizing' in their students. Moreover, from the sum of physical movements that teachers defined as meaningful physical expressions, the resulting numerical values indicate only those physical movements that were observable by the measurers. In short, the figures indicate changes in the students' physical expression that would probably have been perceived by any person present in the classroom[18]. Indeed, these two constraints embody the meaning of the word 'ethno'.

What, then, would the future hold if these two constraints were removed? At the end of this paper, I would like to develop a Science Fiction of Education (SFE) around this question.

First of all, what happens if indicators are set by someone other than those involved in the lesson (teachers and students)? This means that the authority for class evaluation shifts from the teachers themselves, as the class practitioners, to someone outside the classroom. As a result, in order to elevate the evaluation outcomes of their own lessons, teachers would aim for as many students as possible to simultaneously display physical movements that are purported to be expressions of a targeted mental state (e.g., 'activeness', 'spontaneity' or 'personalization and empathizing'). Likely achieved by practitioners performing signalling behaviours that would prompt the desired reaction from students. The result would be a strange classroom landscape, akin to that of a mass game.

Secondly, let us go beyond measuring physical expressions as intersubjectively

[18] Just by someone being there, certain physical movements will already begin to take on symbolic meaning as physical expressions.

expressed/interpreted symbols (which are in essence the impressions that the participants in the interaction have of each other's physical movements). What happens if we seek a more objective and precise measurement of the reality of students' specific physical movements and their corresponding mental states? First of all, the number of cameras will be increased so that the physical movements of individual students can be observed from a closer position, and then AI may take on the measurement so that measurer biases can be avoided. As we ultimately want to understand students' mental state, we may eventually stop being concerned with physical movements that are shaped by cultural biases and individual performance ability, and try to directly capture various physiological changes using sensors embedded in wearable devices. Furthermore, if physiological information is recorded and stored over a long period of time, it will be possible to trace changes in students' mental states and to predict changes to some extent. More precise measures to capture students' state of mind may lead to ways such as the implantation of a chip under the skin. Although students and parents will be able to refuse to do so, they are likely to concede on the grounds that without the chip implantation they would not be able to receive effective teaching. They would be led to make up 'their own mind' (?) and accept this new method[19]. If more people are implanted with these chips, and if they also incorporate GPS functionality, it will be possible to trace students' learning histories quite precisely. If such devices could capture when, for how long, and with whom a student was in contact with, and if such information were available in the form of big data (because the information stored in such devices is not forgotten, it will be instantly recalled at any time and can easily be compared with other similar cases), it would be possible to understand the precise origin of students' state of mind and to predict future changes with a much higher probability. When such an era arrives, such questions as "Who am I?" or "What should I choose at this crossroads?" will be answered far more reliably by Siri or Google than by 'Me' – but then 'I', as a being, will already be regarded as a collection of coincidences. However, by then, there may be no one to even question the meaning of 'My' existence or the meaning of 'My' life anymore.

But why then do we want such precise measurements? Perhaps it is because we want to make as accurate a prediction as possible about the future, which is essentially unknowable. And what is the purpose of such predictions? In a word, it is to avoid death. 'Death' here is symbolic and refers to crisis situations such as unemployment, failure in examinations, 'drop-out', truancy or withdrawal from school. Situations in which the social relationships and flow of life up to that

[19] In terms of health services, it is easy to imagine a situation in which obtaining favourable health and life insurance would depend on undergoing a chip implantation operation. The concept of 'one's own mind' in this case will no longer be distinguishable from others, as is the case with brainwashing.

point are severed. In the medical world – if quality of life is not questioned – the death of the patient may be avoided in both a literal and symbolic sense. In education, however, the situation is quite different. The discontinuity of such a life is ambivalent, and, despite the risk of failure, it can also be an opportunity for transcendence, leaps, adventures, and challenges through *a posteriori* assignment of meaning [Lenzen, D. 1993]. In other words, in the case of education, avoiding 'death' through precise measurement also means extinguishing these opportunities.

When considering the above science fictions, the two constrains in the ethnometric method do not mean that the method is incomplete, or that it remains on the way towards the scientification of educational research. As G. Biesta states, education is "moral practice" more than a "technological activity" and, therefore, if the question for teachers is not just "what works" but also "what is appropriate for these children in this situation" [Biesta, G. J. 2020, pp. 57-58], then at least the above two constraints may be inevitable for educational practice and theory.

Furthermore, Biesta points to the problem of the "separation of means and ends of professional conduct" in evidence-based education. According to Biesta, means and ends are "internally or constitutively" linked in education [Ibid, p. 54 and 57] and therefore, in the search for "good education", a "deliberative democracy" [Ibid, p. 143] around the appropriate means and ends of education is needed. I would like to conclude this paper by relating this point by Biesta to this research and suggesting the possibility of deliberative democracy. I would like to refer again to indicator 1 'nodding' and indicator 7 'various neck movements such as nodding, shaking one's head from side to side, and tilting the head sideways appear in succession'. Indeed, in this research, the frequency of appearance/observation of these two indicators was never high. It is rather important however, that by settling on both indicators despite their similarity (both essentially referring to neck and head movements), the two class practitioners left some room for further discussion. In fact, these two indicators suggest contradictions or fluctuations in teachers' ideas about what is 'personalization and empathizing'. That is, whereas indicator 1 can be associated with the acceptance of existing modes of perception and values, indicator 7 can also be related to their questioning or rejection. With this in mind, it may be possible to internally and constructively link the step of indicator-setting to the deliberations around the lessons' goal – specifically, the meaning of the state of 'personalization and empathizing' as a teaching goal. In any case, even if it is not possible to stop the bus of evidence-based approaches with our bare hands, it is possible to get on the bus and operate the brakes and steering wheel.

Reference

Biester, G. J. (2020). *Yoi Kyouiku towa Nani ka, Rinri, Seiji, Minshushugi (Good Education in an*

Age of Measurement, Ethics, Politics, Democracy), translated by H. Fujii & H. Tamaki. Hakutakusha.

Campbell, D. T. (1975). Assessing the impact of planned social change. In G. M. Lyons (Ed.). *Social Research and Public Policies: The Dartmouth/OECD Conference*. Hanover, New Hampshire: Dartmouth College, The Public Affairs Center, pp. 3-45.

Fujikawa, N. (2020). Taikenkatsudo no Chosa to Bunseki (Survey and analysis of experiential activities). In Japan Contemporary Dance Network (JDDN), NPO (Ed.): Heisei 30 Nendo, Monbukagakusho Seishonen no Taikenkatsudo no Suishin 'Jikokoteikan Kojyo Project': Sozoteki na Dance o mochiita Jido-Seishonen no Jikokoteikan Kojyo Project (FY 2008, Ministry of Education, Culture, Sports, Science and Technology, Promotion of Experiential Activities for Youth 'Self-affirmation Improvement Project': Project to improve children and youth's sense of self-affirmation, using creative dance)", pp. 22-29.

Gebauer, G. (2017). Goffman no Engekiteki Koui no Riron kara Bourdieu no Habitus Gainen e (From Goffman's Theory of Theatrical Action to Bourdieu's Concept of Habitus, translated by N. Fujikawa. In N. Fujikawa (Ed.). *Jinsei no Choritsushitachi: Doteki Dramaturgy no Tenkai (The Tuners of Life: The Development of Dynamic Dramaturgy).* Shunposha, pp. 73-88.

Lenzen, D. (1993). *Krankheit als Erfindung: Eingriffe in die Kultur.* FISCHER Taschenbuch.

Lind, G. (2011). Verbesserung des Unterrichts durch Selbstevaluation: Ein Plädoyer für unverzerrte Evidenz. In J. Bellmann & Th. Müller (Hrsg.). *Wissen, was wirkt: Kritik evidenzbasierter Pädagogik*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 173-195.

Otsuka, T. & Shiotsu, H. (2022). 'Tokubetsu no Kyoka Dotoku' niokeru Gakushushido Katei ni kansuru Kenkyu: Shogakko niokeru Jugyojissen no Kensho o tsujite (A Study on the Learning Guidance Process in 'Special Subject Moral Education': Through Verification of Teaching Practice in Elementary School). In Shimane University Graduate School of Teaching (Ed.). *Research on School Education Practice*, Vol. 5, pp. 23-42.

Sackett, D. L. et al. (2000). *Evidence-Based Medicine*, Second Edition. Churchill Livingstone (Japanese translation published as Sackett, D. L. et al. (2003). *Evidence-Based Medicine*, Elsevier Science).

Shibata, Y. (2021). Hyojyobunseki no Katsuyo niyoru Jugyo Bunseki no Kokoromi: Kyoshi no Hatsuwanaiyo niyoru Kodomo no Kanjyo no Henka ni chakumokushite (A Trial of Lesson Analysis Using Facial Expression Analysis: Focusing on the changes in children's emotions caused by the content of the teacher's utterance). In *Research Reports of the Japan Society for Educational Technology*, No. 4, pp. 216-222.

Table 4: Indicators' selection rationale

Indicator no.	Concrete physical expression	Reason for selecting the indicator
1	Nodding (to teacher's statements)	After watching the video, it was noticed that the number of scenes of students' 'nodding' was less than the class practitioners had expected. Overall, there were not many frequent nodding scenes when listening to statements made by the teacher or friends. When looking at the actual situations in which students nodded, it was found that nodding was seen in situations of intense agreement with what others said ("yes, yes!")' or 'when students could imagine a concrete situation ("yes!")'.
2	Leaning forward (excluding cases involving raising hands)	After watching the video, it was also noticed that children do not lean forward to listen to what is being said unless there is something wrong. At the stage before they lean forward, students think about the question in their own way, either they have own ideas or they have tried to think about it in their own way, but they are not able to determine what to think (they feel confused). In such situations, they may become very curious about what others say or want to compare it with their own ideas, which may make them lean forward.
3	Widening eyes (lively gaze)	When students' eyes widen or there is an obvious change in their facial expression, we believe that some kind of internal value fluctuation has occurred. Such a change is likely to occur when listening to a friend's statement. After listening to a statement, empathic feelings like "Yes, that's true!" and "I have had similar experiences and feelings!", as well as feelings of surprise at obvious differences between their own ideas and those of the other person like "I didn't know there was such a way of thinking!" and "Really?" may appear around their eyes and in their facial expression.
4	Upward gaze (a short	We think that when students look up after hearing the teacher's question it is a sign that they are seriously considering the question in their own way. It is difficult for children to immediately

	pause in response to a question and staring into space)	determine how and what to think about questions that do not have a correct answer, that is, questions that require they think about their own experiences. We think that the way their eyes are looking upwards shows that they are remembering what they themselves have experienced in the past and thinking more concretely about the question "What if it were me?"
5	Fixed gaze (gaze stops moving in response to the question posed by teacher)	"When someone speaks, look at them carefully and listen thoughtfully" is something that is taught in various situations at school, but in practice it is difficult to ensure this way of listening every time. It is practically unavoidable that students may be writing notes, looking away or misbehaving due to a break in concentration. In these situations, there are sometimes moments when students' eyes are drawn towards the teacher in response to the teacher's question, and they stare firmly at the person who is speaking, without their eyes darting about. We believe that this is a sign that the students are trying to stop and think about the question in their own way.
6	Conversation starts 'right after' encouraging discussion with neighbouring students in the course of the lesson	Dialogue may be held in pairs or groups with a neighbouring or nearby friend. If the students do not know what to discuss in response to the prompt "Talk to someone nearby", if they do not find a personal connection to the question, the pair/group conversation will not be active. Sometimes they gather near each other for the time being, but they are silent. Sometimes, however, as soon as the teacher encourages dialogue in pairs/groups, the students immediately turn their bodies towards their friends and start talking to each other. This may be an expression of students' desire to solve the problem as they already have some ideas of their own in response to the teacher's prompting, or because it is a difficult subject to think about as related to themselves and they want to talk it through with their friends.

7	Various neck movements such as nodding, shaking one's head from side to side and tilting the head sideways appear in succession	Sometimes you would see students in class shaking their heads quickly and in small increments. The act of 'shaking one's head' is thought to be an expression of a negative opinion or an opposite (different) opinion in response to what the teacher has said or what a friend has said. However, even if a student had such thoughts, if they do not relate personally to the teacher's question, there will be no noticeable change in students' neck movements. This kind of change can be seen when students are thinking about an issue in their own way, taking the matter seriously and have their own ideas firmly in mind. When various movements are observed, such as tilting the head to the side as well as shaking the head, we believe that the students are in a state of intense conflict about values. The state of conflict and hesitation (*moyamoya*) after thinking through the teacher's question based on their own understandings such as 'understanding of value', 'understanding of human beings' and 'understanding of others' may be expressed as such movements.

Class session	Dndicator no.	Normal acurracy															
		Camera 1						Camera 2									
		Number of samples *n1 /n2*	Unweighted			Weighted			Number of samples *n1 /n2*	Unweighted			Weighted				
			Two-sided P-value	Median change	Change in average ranking	Two-sided P-value	Median change	Change in average ranking		Two-sided P-value	Median change	Change in average ranking	Two-sided P-value	Median change	Change in average ranking
3	1	50/25	non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00	50/25	non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00
	2		0.519	0.00→0.00	38.46→37.08	0.519	0.00→0.00	38.46→37.08		(-)0.041*	0.00→0.00	39.74→34.52	(-)0.041*	0.00→0.00	39.74→34.52
	3		0.153	0.00→0.00	38.50→37.00	0.153	0.00→0.00	38.50→37.00		0.153	0.00→0.00	38.50→37.00	0.153	0.00→0.00	38.50→37.00
	4		0.077	0.00→0.00	38.75→36.50	0.077	0.00→0.00	38.75→36.50		0.035**	0.00→0.00	35.79→42.42	0.014**	0.00→0.00	35.50→43.00
	5		(-)0.000***	0.10→0.05	44.75→24.50	(-)0.000***	0.11→0.05	44.70→24.60		0.061	0.08→0.05	41.35→31.30	(-)0.050*	0.08→0.05	41.47→31.06
	6		(-)0.039*	0.00→0.00	39.00→36.00	(-)0.039*	0.00→0.00	39.00→36.00		non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00
	7		(-)0.020*	0.00→0.00	39.25→35.50	(-)0.020*	0.00→0.00	39.25→35.50		0.172	0.00→0.00	39.31→35.38	0.253	0.00→0.00	39.15→35.70
5	1	50/50	0.100	0.00→0.00	48.55→52.45	0.100	0.00→0.00	48.55→52.45	50/50	0.123	0.00→0.00	52.59→48.41	0.123	0.00→0.00	52.59→48.41
	2		0.000***	0.00→0.04	38.43→62.57	0.000***	0.00→0.04	38.43→62.57		0.178	0.00→0.00	47.17→53.83	0.160	0.00→0.00	47.02→53.98
	3		0.118	0.00→0.00	48.38→52.62	0.118	0.00→0.00	48.38→52.62		0.501	0.00→0.00	51.72→49.28	0.501	0.00→0.00	51.72→49.28
	4		0.317	0.00→0.11	51.00→50.00	0.317	0.00→0.00	51.00→50.00		0.621	0.00→0.00	49.40→51.60	0.732	0.00→0.00	49.74→51.26
	5		0.198	0.19→0.11	54.39→46.61	0.128	0.21→0.12	55.03→45.97		0.071	0.22→0.17	55.71→45.29	0.074	0.22→0.17	55.61→45.39
	6		0.000***	0.00→0.00	45.00→56.00	0.000***	0.00→0.00	45.00→56.00		0.000***	0.00→0.00	44.41→56.59	0.000***	0.00→0.00	44.41→56.59
	7		0.403	0.00→0.00	49.50→51.50	0.423	0.00→0.00	49.545→51.46		0.859	0.00→0.00	50.13→50.87	0.851	0.00→0.00	50.11→50.89
6	1	50/31	0.317	0.00→0.00	41.31→40.50	0.317	0.00→0.00	41.31→40.50	50/30	0.542	0.00→0.00	40.93→39.78	0.542	0.00→0.00	40.93→39.78
	2		(-)0.000***	0.07→0.00	50.97→24.92	(-)0.000***	0.07→0.00	51.03→24.82		(-)0.000***	0.08→0.01	49.28→25.87	(-)0.000***	0.08→0.01	49.22→25.97
	3		(-)0.010***	0.00→0.00	42.86→38.00	(-)0.010***	0.00→0.00	42.86→38.00		0.720	0.00→0.00	40.29→40.85	0.720	0.00→0.00	40.29→40.85
	4		0.088	0.00→0.00	39.29→43.76	0.177	0.00→0.00	39.79→42.95		0.533	0.00→0.00	39.33→42.45	0.315	0.00→0.00	38.81→43.32
	5		(-)0.010***	0.56→0.20	49.43→27.40	(-)0.000***	0.72→0.23	49.66→27.03		(-)0.010**	0.61→0.14	50.82→23.30	(-)0.000***	0.64→0.14	51.45→22.25
	6		non-computable	0.00→0.00	41.00→41.00	non-computable	0.00→0.00	41.00→41.00		non-computable	0.00→0.00	40.00→40.00	non-computable	0.00→0.00	40.50→40.50
	7		0.097	0.00→0.00	38.95→44.31	0.097	0.00→0.00	38.95→44.31		0.335	0.00→0.00	38.63→43.62	0.360	0.00→0.00	38.73→43.45
7	1	50/27	non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00	50/27	non-computable	0.00→0.00	25.50→64.00	0.317	0.00→0.00	39.27→38.50
	2		0.601	0.00→0.00	39.74→37.63	0.601	0.00→0.00	39.74→37.63		0.689	0.00→0.00	38.43→40.06	0.171	0.00→0.00	36.94→42.82
	3		0.077	0.00→0.00	39.81→37.50	0.078	0.00→0.00	39.81→37.50		0.133	0.00→0.00	40.18→36.82	0.010***	0.00→0.00	35.22→46.00
	4		0.317	0.00→0.00	38.50→39.93	0.317	0.00→0.00	38.50→39.93		0.070	0.00→0.00	41.30→34.74	0.051	0.00→0.00	35.59→45.32
	5		0.010***	0.03→0.12	31.54→52.82	0.010***	0.03→0.13	31.14→53.56		0.966	0.03→0.03	39.08→38.85	0.046**	0.03→0.12	35.21→46.02
	6		non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00		non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00
	7		0.317	0.00→0.00	39.27→38.50	0.317	0.00→0.00	39.27→38.50		0.000***	0.00→0.03	43.86→30.00	0.041*	0.00→0.03	35.41→45.65

Table 5: Change in p-value, median and mean rank by class sessions and indicators (Normal accuracy)

Class session	Indicator no.	High acurracy													
		Camera 1						Camera 2							
		Number of samples *n1 /n2*	Unweighted			Weighted			Number of samples *n1 /n2*	Unweighted			Weighted		
			Two-sided P-value	Median change	Change in average ranking	Two-sided P-value	Median change	Change in average ranking		Two-sided P-value	Median change	Change in average ranking	Two-sided P-value	Median change	Change in average ranking
3	1	50/25	non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00	50/25	non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00
	2		0.772	0.00→0.00	37.71→38.58	0.786	0.00→0.00	37.73→38.54		0.317	0.00→0.00	38.25→37.50	0.317	0.00→0.00	38.25→37.50
	3		(-)0.039**	0.00→0.00	39.00→36.00	(-)0.039**	0.00→0.00	39.00→36.00		non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00
	4		0.623	0.00→0.00	38.77→36.46	0.623	0.00→0.00	38.77→36.46		0.002***	0.00→0.00	34.50→45.00	0.002***	0.00→0.00	34.50→45.00
	5		(-)0.006***	0.05→0.00	41.91→30.18	(-)0.006***	0.07→0.00	41.94→30.12		0.050	0.00→0.00	40.58→32.84	0.050	0.00→0.00	40.58→32.84
	6		non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00		non-computable	0.00→0.00	38.00→38.00	non-computable	0.00→0.00	38.00→38.00
	7		0.844	0.00→0.00	37.78→38.44	0.844	0.00→0.00	37.78→38.44		0.665	0.00→0.00	37.76→38.48	0.317	0.00→0.00	38.25→37.50
5	1	50/50	non-computable	0.00→0.00	50.50→50.50	non-computable	0.00→0.00	50.50→50.50	50/50	non-computable	0.00→0.00	50.50→50.50	non-computable	0.00→0.00	50.50→50.50
	2		0.284	0.00→0.00	53.10→47.90	0.230	0.00→0.00	53.41→47.59		0.561	0.00→0.00	51.00→50.00	0.561	0.00→0.00	51.00→50.00
	3		0.133	0.00→0.00	47.81→53.19	0.167	0.00→0.00	48.02→52.98		0.317	0.00→0.00	51.00→50.00	0.317	0.00→0.00	51.00→50.00
	4		0.516	0.10→0.06	52.43→48.57	0.482	0.11→0.07	52.58→48.42		0.989	0.00→0.00	50.49→50.51	0.989	0.00→0.00	50.49→50.51
	5		0.457	0.17→0.10	52.69→48.31	0.347	0.18→0.10	53.25→47.75		0.862	0.34→0.34	51.02→49.98	0.680	0.36→0.32	51.73→49.27
	6		0.000***	0.00→0.00	44.50→56.50	0.000***	0.00→0.00	44.50→56.50		0.000***	0.00→0.00	44.50→56.50	0.000***	0.00→0.00	44.50→56.50
	7		0.270	0.00→0.00	52.93→48.07	0.245	0.00→0.00	53.06→47.94		0.989	0.00→0.00	50.51→50.49	0.989	0.00→0.00	50.51→50.49
7	1	50/31	0.317	0.00→0.00	41.31→40.50	0.317	0.00→0.00	41.31→40.50	50/30	non-computable	0.00→0.00	40.50→40.50	non-computable	0.00→0.00	40.50→40.50
	2		(-)0.000***	0.06→0.00	47.16→31.07	(-)0.000***	0.06→0.00	47.04→31.26		(-)0.000***	0.01→0.00	46.68→30.20	(-)0.000***	0.05→0.00	46.15→31.08
	3		(-)0.000***	0.00→0.00	45.50→33.74	(-)0.000***	0.00→0.00	45.54→33.68		0.202	0.00→0.00	41.53→38.78	0.202	0.00→0.00	41.53→38.78
	4		(-)0.000***	0.33→0.13	50.48→25.71	(-)0.000***	0.42→0.16	50.56→25.58		0.434	0.00→0.00	41.46→38.90	0.434	0.00→0.00	41.46→38.90
	5		(-)0.000***	0.38→0.17	49.07→27.98	(-)0.000***	0.85→0.33	49.30→27.61		(-)0.000***	0.73→0.30	47.58→28.70	(-)0.000***	0.00→0.00	47.74→28.43
	6		0.15	0.00→0.00	40.00→42.61	0.150	0.00→0.00	40.00→42.61		non-computable	0.00→0.00	40.50→40.50	non-computable	0.00→0.00	40.50→40.50
	7		0.290	0.00→0.00	38.58→43.53	0.446	0.00→0.00	39.63→43.21		0.483	0.00→0.00	39.72→41.80	0.472	0.00→0.00	39.70→41.83
7	1	50/27	non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	38.00→38.00	50/27	non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	38.00→38.00
	2		0.277	0.00→0.00	40.65→35.94	0.280	0.00→0.00	40.64→35.96		0.298	0.00→0.00	37.93→40.98	0.298	0.00→0.00	37.93→40.98
	3		(-)0.025**	0.00→0.00	40.98→35.33	(-)0.025**	0.00→0.00	40.98→35.33		0.153	0.00→0.00	39.54→38.00	0.153	0.00→0.00	39.54→38.00
	4		0.009***	0.00→0.06	34.67→47.02	0.009***	0.00→0.00	34.70→46.96		(-)0.043**	0.00→0.00	40.86→35.56	(-)0.043**	0.00→0.00	40.86→35.56
	5		0.524	0.00→0.00	37.96→40.93	0.574	0.00→0.04	38.09→40.69		0.000***	0.00→0.17	28.22→58.96	0.000***	0.00→0.04	28.30→58.82
	6		non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00		non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00
	7		0.083	0.00→0.00	36.46→43.70	0.079	0.00→0.00	36.42→43.78		non-computable	0.00→0.00	39.00→39.00	non-computable	0.00→0.00	39.00→39.00

Table 6: Change in p-value, median and mean rank by class sessions and indicators (High accuracy)