

Characteristics and Cognitive Processes of Teachers' Test Scoring in Japanese Elementary and Secondary Education

Satoshi Nakagawa ^{*}, Ryo Saito [†], Shota Itagaki [†],
Tomohiro Yamamoto [‡], Kazunori Sato [§], Tatsuya Horita ^{*}

Abstract

Teachers in elementary and secondary education in Japan are working long hours and this is a big issue. One of the reasons for the long hours is test scoring. In this study, a survey on different school types (elementary, junior high, and senior high schools) in Japan was conducted to determine the characteristics of test scoring, the cognitive processes involved, and the burden of the test scoring processes (correct/incorrect judgment, score calculation, and score recording). The results showed that all types of school teachers in Japan felt burdened due to test scoring task processes. Additionally, teachers in all school types try to ascertain the degree of students' understanding through the test scoring task processes. A positive correlation was found between the teachers' ascertaining the degree of students' understanding and reflecting on their own teaching. This study revealed that the cognitive processes of scoring (teachers' ascertainment of students' degree understanding, reflection on their own teaching, and sense of burden) are slightly different depending on the school type. However, teachers in all school types regarded test scoring as an educational activity and not merely a simple task. This should be borne in mind, whether the test scoring is done manually or by a computer system.

Keywords: test scoring, cognitive process, students' understanding, reflection on teaching, elementary and secondary education

1 Introduction

In every country, school education is an essential social system, and Japan is no exception. Elementary and secondary education in Japan comprises elementary school, junior high school, and senior high school. The duration of compulsory education is nine years, divided into six years of elementary school (7–12 years of age) and three years of junior high school (12–15 years of age) [1]. Senior high school (age 16 onwards) is not compulsory but the

^{*} Tohoku University, Miyagi, Japan

[†] Miyagi University of Education, Miyagi, Japan

[‡] Kagoshima University, Kagoshima, Japan

[§] Shinshu University, Nagano, Japan

rate of advancement from junior high school as of 2018 is more than 98% [2]. The maximum number of students per class in elementary school and junior high school is 40, which is also the standard number of students per class in senior high school [3]. The teacher assignment system for elementary schools is mainly class-based, with subject-based teachers only for certain grades and subjects (e.g., science, social studies, and music). Meanwhile, junior and senior high school teachers' assignments are all subject-based [4].

The tasks of teachers in Japan are vastly different from those in other countries. Whereas in other countries, teachers' tasks are specific to their lessons, teachers' tasks in Japan include not only lessons on the subjects being taught but also include integrated moral education (to cultivate a rich sense of humanity) and physical education (to cultivate health and physical strength). Teachers in Japan aim to multilaterally cultivate the students' qualities and capacities for their future lives [5]. This is referred to as the "Japanese Model of School Education," which has received a positive evaluation internationally [6]. This has manifested in the academic ability of Japanese students, who consistently achieved high scores among the 37 Organization for Economic Co-operation and Development (OECD) member nations in the 2018 Program for International Student Assessment (PISA), ranking first in mathematics, second in science, and eleventh in reading [7].

However, the working hours of teachers in Japan are the longest among OECD member countries and this is a big problem [8]. Previous reports showed that long working hours make teachers suffer emotional strain and cause health issues [9][10]. Furthermore, long working hours are associated with the risk of inducing the desire to quit [11]. In 2018, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan surveyed the reasons for the long working hours of teachers [12]. The results indicated the following:

- The most time-consuming tasks in elementary schools, except for lessons and lesson preparations, are "morning tasks (35 min)" and "grade processing (33 min)."
- The most time-consuming tasks in junior high schools, except for lessons and lesson preparations, are "club activities (41 min)" and "grade processing (38 min)."

In Japanese elementary schools and junior high schools, the common task that takes up time is grade processing. Furthermore, according to a follow-up survey conducted by MEXT in elementary schools and junior high schools, teachers who work 60 hours or more per week devote more time to grade processing than teachers who work no more than 60 hours per week. MEXT concluded that grade processing is the main factor that causes long working hours [13].

The content of grade processing in schools was defined by MEXT as tasks related to grade processing, test question production, scoring/evaluation, report card entry, school report production, cumulative guidance, student record production, and so on [12]. Elementary school teachers use "unit test" scores and junior high school teachers use "periodical test" scores for creating the students' report cards [14]. Several scoring support systems have been developed in Japan to reduce teachers' excessive workload [15][16][17]. MEXT has also published a case study demonstrating how a scoring support system can decrease test scoring time [18]. This case study showed that at a school that implemented a scoring support system, teachers spent 25 fewer hours on correct and incorrect decisions per year, 108 fewer hours on score calculations, and 10.8 fewer hours on posting scoring results to a ledger [18].

Scoring is a demanding and high-priority task for teachers because it is highly related to grade processing. It has an educational aspect, that is, understanding and evaluating students and reflecting on their own teaching. Kajita [19] identified two aspects of test scoring: simple tasks and educational activities. He states the following as a formative assessment: “Achievement tests make it possible to provide an appropriate supplementary instruction for students because the tests give teachers information about children’ understanding and not understanding [19].”

However, several studies have reported that teachers find it difficult to determine students’ level of understanding through a scoring support system [20][21]. We should not disregard the fact that teachers ascertain the students’ level of understanding through test scoring in order to understand the characteristics and cognitive processes. In this study, the characteristics of test scoring and cognitive processes—ascertaining degree of students’ understanding, reflection on their own teaching, and sense of burden—are investigated.

2 Previous Study

2.1 Test Characteristics in Each Type of School

Nakagawa et al. [17] surveyed the characteristics (i.e., number of questions, number of tests in a year, and test takers per test) of the tests used for evaluation in elementary, junior high schools and senior high schools, based on prior studies [14][22][23][24]. They are compiled in Table 1. The following were observed:

In elementary schools,

- commercially available unit tests are used for evaluation,
- the number of tests in a year is greater than that of junior and senior high schools, and
- the number of questions and test takers per test are relatively small.

In junior and senior high schools,

- periodical tests that are self-produced by teachers are used for evaluation,
- the number of tests in a year is less than that of elementary schools, and
- the number of questions and test takers per test are relatively large.

Table 1: Characteristics of Tests in Each Type of School [17]

Type of school	Scorer (Subject scored)	Tests linked to grade processing	Tests used	Tests/year	Questions/test	Test takers/scorer (classes tested)
Junior/senior high school	Subject teacher (subjects taught)	Periodical tests	Teacher- created test	Approx. 5	Approx. 60	Approx. 120 (multiple classes)
Elementary school	Classroom teacher (multiple subjects)	Unit tests	Commercially available test	Approx. 60	Approx. 25	Approx. 30 (1 class)

Note: Compiled based on interviews with the Benesse Educational Research and Development Institute [14][23][24], Systems Research & Development Institute of Japan [22]

Nakagawa et al. [17] arranged the flow of test scoring by dividing them into three categories by referencing prior studies [25][26]. The first step is “preparation” (i.e., producing and duplicating test questions). The second step is “implementation” (i.e., distributing, supervising, and collecting tests). The final step is “scoring” (i.e., correct/incorrect judgment, score calculating, and recording). The individual steps are defined as follows and illustrated in Figure 1:

- Correct/incorrect judgment: Judging whether the answers are correct or incorrect,
- Score calculating: Calculating scores of correct answers and writing them down on answer sheets,
- Score recording: Recording (inputting) the scores of each student in the register system/teacher’s school work support system.

Note: Compiled with reference to Japan Association for Research on Testing [25] and Study Group on the Use of School Educational Materials [26]

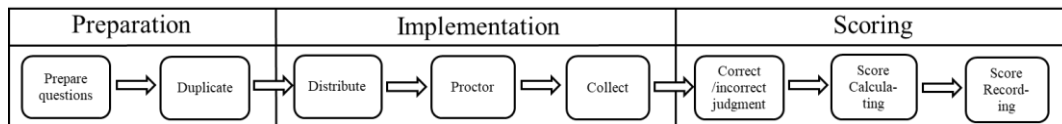


Figure 1: Test Procedure [17]

2.2 Classification of Correct/Incorrect Judgment Method in Scoring

Nakagawa et al. [17] observed and surveyed test scoring in the manual processes (i.e., pen-and-paper), especially in correct/incorrect judgments, and classified the method of correct/incorrect judgments as follows:

- Question-based judgment method (Question Method): A teacher makes the correct/incorrect judgments for each question for all test takers.
- Test taker-based judgment method (Student Method): A teacher makes the correct/incorrect judgments for all the answers from each test taker.
- Block-based judgment method (Block Method): A teacher makes the correct/incorrect judgments for blocks comprising viewpoints or major questions.

Nakagawa and Horita [27] conducted semi-structured interviews on junior high school teachers in line with the details of the scoring methods as defined by Nakagawa et al [17], to survey what they attach importance to during each task process of the scoring. Typification was performed on the results of each interview, with regard to the cognitive processes of teachers in each task process of the scoring, and the question items in the questionnaire survey were created.

3 Survey

3.1 Method

A mark-sheet type questionnaire (Survey of teachers regarding test scoring) was created based on Nakagawa and Horita [27]. The questionnaire was composed of questions about

anonymous basic information of the test takers and the characteristics and cognitive processes in test scoring. Regarding the demographic information, we asked about the type of school, type of test used primarily to evaluate grades, frequency of implementation, test producer, number of questions, and number of test takers. On the characteristics and cognitive processes in test scoring, we asked about the correct/incorrect judgment method, sense of burden, importance placed on the degree of students' understanding as well as a reflection on their own teaching in each work process. Table 2 shows the questions related to this study.

Table 2: Question Details

Questionnaire content		Answer method	Options
Type of school	Type of school you belong to/work at.	Select only one	Elementary school (Home room) / Elementary school (Subject-based) / Junior high school/Senior high school / Other
Type of test	Type of test you emphasize the most for results evaluation	Select only one	Short test (individual lessons) / Unit test (individual unit) / Periodical test (1-2 times per term) / Grade test (test of actual ability) / Other
Producer	Producer of the tests that you score	Select only one	Yourself / Divided between multiple teachers / Other teacher (you provided advice) / Manufacturer / Other
Number of test takers	Test takers of the tests that you score	Select only one	1-20/21-40/41-60/61-80/81-100/101-120/121 or more
Correct/incorrect judgement	The main correct/incorrect judgment method you use when scoring	Select only one	By question (compile only the same questions for all test takers) / By block (units of blocks compiled from each main question or by viewpoint, etc.) / By test taker (compile all answers by each test taker) / Other
Burden	Do you feel burdened during correct/incorrect judgments	Select only one	1. I feel a strong sense of burden. 2. I feel some sense of burden. 3. Neutral. 4. I don't have a sense of burden. 5. I feel no sense of burden whatsoever.
	Do you feel burdened during score calculating		
	Do you feel burdened during score recording		
Ascertaining degree of students' understanding	Do you give importance to correct/incorrect judgments	Select only one	1. No importance at all. 2. Very little importance. 3. Neutral. 4. Some importance. 5. Great importance
	Do you give importance to calculating scores		
	Do you give importance to recording scores		
Reflecting on teachers' own teaching	Do you give importance to correct/incorrect judgments	Select only one	1. No importance at all. 2. Very little importance. 3. Neutral. 4. Some importance. 5. Great importance
	Do you give importance to calculating scores		
	Do you give importance to recording scores		

3.2 Participants and Survey Period

The participants comprised 842 teachers working in 50 schools, including 32 public elementary schools, 2 private elementary schools, 1 elementary school affiliated with a national university, 8 public junior high schools, 3 private junior high schools, 1 junior high school affiliated with a national university, 1 public senior high school, and 2 private senior high schools. The questionnaire was sent to the participants in late March 2020, and

responses were received by early April 2020.

We excluded 142 participants who gave partially incomplete responses or multiple responses to single-choice questions. We also excluded 17 participants that marked “Other” for type of school (7 participants from 3 public elementary schools, 9 participants from 2 public compulsory education schools, and 1 participant from a private junior high school). The ratio of the types of schools for the 683 participants is shown in Figure 2.

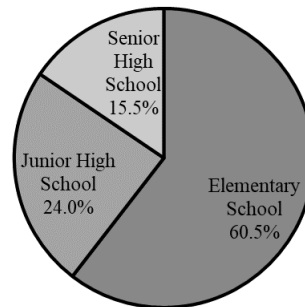


Figure 2: Ratio of Types of Schools ($n = 683$)

4 Results

4.1 Characteristics of Tests Used to Evaluate Students' Grades in Each Type of School

Figure 3 shows the test used to evaluate the students' grades in each type of school. While elementary school teachers use unit tests for evaluation (91.5%), junior and senior high school teachers use periodical tests (92.7%/95.3%) for evaluation. Figure 4 shows the producers of tests used by teachers. Elementary school teachers use commercially available tests (88.1%) and junior and senior high school teachers use non-commercially available tests (By oneself + Together + Other teachers, 96.9%/97.2%). These results do not contradict prior studies (Table 1). Elementary school teachers are class-based teachers and conduct many tests—implement tests for each unit—and use the commercially available tests. Junior and senior high school teachers are subject-based teachers and many teachers are only responsible for one subject; thus, they conduct tests relatively less frequently. They can afford to produce and control the tests themselves.

Figure 5 shows the correct/incorrect judgment method that teachers use, by type of school. While half the elementary school teachers use the Student Method (61.7%), half the junior and senior high school teachers use the Block Method (49.4%/63.2%). Figure 6 shows the number of test takers (either 40 and below or 41 and above) by type of school. Most elementary school teachers marked 40 test takers and below (85.5%), and most junior high school teachers and all senior high school teachers marked 41 test takers or above (94.5%/100%). As a supplementary note, 14.5% of elementary school teachers chose the option of 41 test takers or more because they are subject-based teachers, while 20% teachers at small-scale junior high schools chose the option of 40 test takers or less, despite being subject-based teachers.

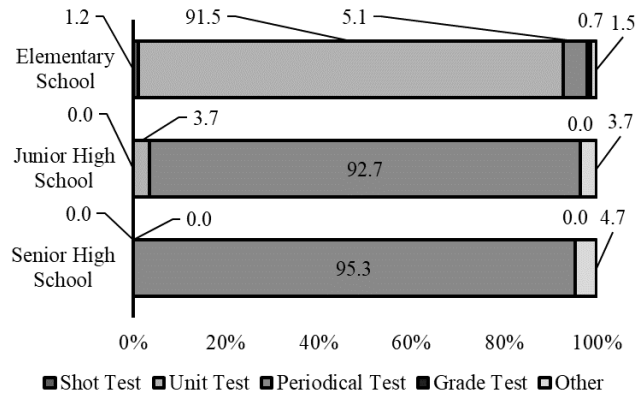


Figure 3: Type of Test (n = 683)

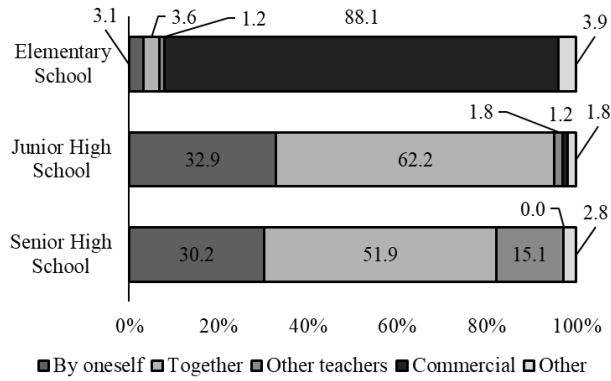


Figure 4: Producer of Test (n = 683)

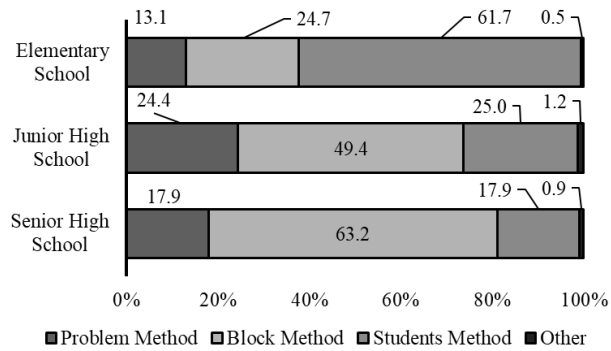
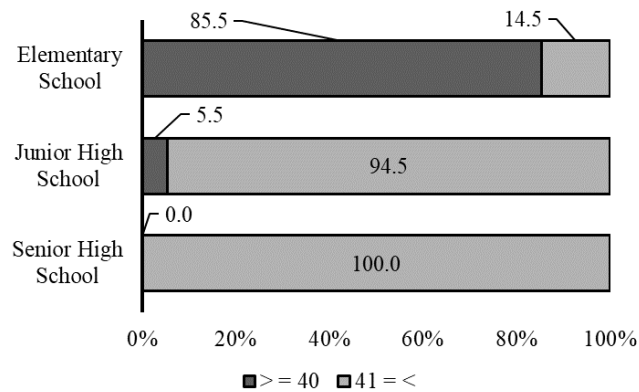


Figure 5: Correct/Incorrect Judgment Method (n = 683)

Figure 6: Number of Test Taker ($n = 683$)

4.2 Sense of Burden

The “sense of burden regarding test work processes” was surveyed. A five-point scale from 1 (I feel a strong sense of burden) to 5 (I feel no sense of burden whatsoever) was used. The scores were reversed and sought by replacing a numerical value between 1 (weaker) and 5 (stronger). The responses to the sense of burden were analyzed using one-sample t -test by school type. The mean value was 3 (i.e., “Neither”) on the five-point scale. While the answer “lower than 3” means “feel no burden,” the answer “higher than 3” means “burden.” The results are shown in Table 3.

It is clear that the teachers in elementary and junior high schools feel burdened in all work processes. High school teachers felt burdened during correct/incorrect judgment and score calculating, but they felt neither “burden” nor “no burden” when score recording.

Table 3: Burden in Test Scoring ($n = 683$)

	Correct/Incorrect					Score Calculating					Score Recording				
	M	(SD)	t	df	p	M	(SD)	t	df	p	M	(SD)	t	df	p
Senior High School	3.84	(1.23)	7.04	105	.000	3.27	(1.31)	2.15	105	.034	2.98	(1.18)	-0.17	105	.869
Junior High School	3.82	(1.16)	9.12	163	.000	3.54	(1.20)	5.73	163	.000	3.32	(1.21)	3.43	163	.001
Elementary School	3.67	(1.11)	12.26	412	.000	3.39	(1.18)	6.65	412	.000	3.51	(1.15)	9.05	412	.000

A comparison of burden was made between groups (types of schools: elementary, junior high, and senior high schools) using ANOVA. In correct/incorrect judgment, the significance was not confirmed ($F(2, 680) = 1.65, p = .194$). In score calculating, the significance was likewise not confirmed ($F(2, 680) = 1.64, p = .194$). In score recording, the significance was confirmed ($F(2, 680) = 9.03, p < .001$). As a result of Tukey’s multiple comparison test, significant differences were found between elementary schools and senior high schools ($p < .001$) and between junior high schools and senior high schools ($p < .05$).

4.3 Ascertaining the Degree of Students' Understanding

The factor of “ascertaining the degree of students’ understanding in each of the scoring task processes” was surveyed. This question used a five-point scale from 1 (No importance at all) to 5 (Great importance). The scores were sought by replacing the answer results with a numerical value between 1 (weaker) and 5 (stronger). The responses to ascertain the degree of students’ understanding were analyzed using a one-sample *t*-test by school type. The results are shown in Table 4. It is clear that the teachers attach importance to all the items in all the work processes in all the school types.

Table 4: Ascertaining the Degree of Students’ Understanding in Test Scoring (*n* = 683)

	Correct/Incorrect					Score Calculating					Score Recording				
	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>
Senior High School	4.32	(0.74)	18.44	105	.000	4.12	(0.84)	13.82	105	.000	3.58	(1.14)	5.21	105	.000
Junior High School	4.29	(0.72)	22.75	163	.000	4.26	(0.78)	20.62	163	.000	3.75	(1.07)	8.97	163	.000
Elementary School	4.60	(0.67)	48.17	412	.000	4.47	(0.80)	37.34	412	.000	3.92	(1.15)	16.35	412	.000

A comparison of ascertaining the degree of students’ understanding the three types was made between groups (types of schools: elementary, junior high, and senior high schools) using ANOVA. In correct/incorrect judgment, the significance was confirmed ($F(2, 680) = 15.03, p = .000$). As a result of the multiple comparison test, significant differences were confirmed between elementary schools and junior high schools ($p = .000$), and between elementary schools and senior high schools ($p = .002$).

In score calculating, the significance was confirmed ($F(2, 680) = 9.75, p = .000$). As a result of the multiple comparison test, significant differences were confirmed between elementary schools and junior high schools ($p = .012$), and between elementary schools and senior high schools ($p = .000$).

In score recording, the significance was confirmed ($F(2, 680) = 4.50, p = .011$). As a result of the multiple comparison test, significant differences were confirmed between elementary schools and senior high schools ($p = .021$).

4.4 Reflection on Teachers’ own Teaching

The factor “reflection on their own teaching in scoring” was surveyed in each of the scoring task processes. This question used a five-point scale from 1 (No importance at all) to 5 (Great importance). The scores were sought by replacing the answer results with a numerical value between 1 (weaker) and 5 (stronger). The results are shown in Table 5. It is clear that the teachers attach importance to all the items in all the work processes in all the school types.

Table 5: Reflection on Teachers' own Teaching in Test Scoring ($n = 683$)

	Correct/Incorrect					Score Calculating					Score Recording				
	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>	<i>M</i>	(<i>SD</i>)	<i>t</i>	<i>df</i>	<i>p</i>
Senior High School	3.95	(0.77)	12.69	105	.000	3.77	(0.92)	8.67	105	.000	3.36	(1.07)	3.45	105	.001
Junior High School	4.02	(0.96)	13.68	163	.000	3.96	(0.93)	13.20	163	.000	3.52	(1.10)	6.04	163	.000
Elementary School	4.33	(0.77)	34.84	412	.000	4.24	(0.80)	31.43	412	.000	3.77	(1.13)	13.81	412	.000

A comparison of reflection on teachers' own teaching was made between groups (types of schools: elementary, junior high, and senior high schools) using ANOVA. In correct/incorrect judgment, the significance was confirmed ($F(2, 680) = 13.45, p < .001$). As a result of a multiple comparison test, significant differences were confirmed between elementary schools and junior high schools ($p < .001$), and between elementary schools and senior high schools ($p < .001$).

In score calculating, the significance was confirmed ($F(2, 680) = 15.78, p < .001$). As a result of the multiple comparison test, significant differences were confirmed between elementary schools and junior high schools ($p < .001$), and between elementary schools and senior high schools ($p < .001$).

In score recording, the significance was confirmed ($F(2, 680) = 7.00, p < .001$). As a result of the multiple comparison test, significant differences were confirmed between elementary schools and junior high schools ($p = .043$), and between elementary schools and senior high schools ($p = .002$).

4.5 Reflection Among Cognitive Processes in Each Test Scoring Task

The coefficient omega (ω) of question responses about cognitive processes (ascertaining degree of students' understanding, reflection on their own teaching, and sense of burden) collapsed across test scoring tasks (correct/incorrect, score calculating, score recording) was calculated ($\omega = .667, .763, .847$) and revealed sufficient internal consistency. The mean score in each scoring task (Table 6) was calculated and a correlation analysis was conducted (Table 7).

There is a positive relationship between ascertaining the degree of students' understanding and reflection on their own teaching ($r = .81$). This tendency is relatively stronger in elementary schools compared to that in other schools ($r = .83, r = .79, r = .70$). These results indicate that trying to ascertain the degree of students' understanding is correlated with improving teaching.

A simple linear regression was calculated to predict reflection on their own teaching based on the ascertaining the degree of students' understanding. A significant regression equation was found ($F(1, 681) = 1276.06, p < .001$), with an R^2 of .652. Ascertaining the degree of students' understanding was found to be a predictor of reflection on teachers' own teaching ($B = 0.91, SE_B = 0.03, \beta = 0.81, t = 35.72, p < .001$). Teachers with higher scores in ascertaining degree of students' understanding tended to emphasize reflecting on their own teaching.

Table 6: Mean Scores in Each Scoring Process

	Correct/Incorrect		Score Calculating	
	Understand	Reflection	Understand	Reflection
<i>M</i>	4.48	4.20	4.36	4.10
<i>SD</i>	0.71	0.84	0.81	0.87
	Score Recording		Mean	
	Understand	Reflection	Understand	Reflection
<i>M</i>	3.83	3.64	4.22	3.98
<i>SD</i>	1.13	1.12	0.69	0.78

Burden				
	Correct/ Incorrect	Score Calculating	Score Recording	Mean
<i>M</i>	3.73	3.41	3.39	3.51
<i>SD</i>	1.14	1.21	1.18	1.02

Note: Mean and Standard Deviation ($n = 683$)

5 Conclusion and Future Issues

This study investigated the characteristics and cognitive processes in teachers' test scoring in Japan. We surveyed several characteristics—such as the types of tests, the number of tests in a year, and the number of test takers—in each type of school. Additionally, we analyzed several cognitive processes—such as the sense of burden, ascertaining the degree of students' understanding, and reflection on their own teaching—in each process of test scoring task (correct/incorrect judgment, score calculating, and score recording).

Scoring is an important task that is closely related to grade processing and it is believed that teachers feel burdened by all the processes of this task, except score recording. However, as Kajita [15] pointed out, teachers made efforts to ascertain the degree of students' understanding and conduct reflections on their own teaching. Thus, teachers do not perceive scoring as a mere task that can be easily replaced by computers but rather as an educational activity associated with learning guidance. Furthermore, elementary school teachers placed more importance on ascertaining the degree of students' understanding and reflecting on their own teaching on the tasks of correct/incorrect judgment and score recording than junior and senior high school teachers. A teacher in elementary school is responsible for only a small number of students, thus, the number of answer sheets to be scored is small and it is easier to grasp the understanding of each student. Additionally, elementary school teachers teaching a single class can easily reflect on their teaching through scoring compared to junior and senior high school teachers.

Next, we discuss the correlation between ascertaining the degree of students' understanding and reflecting on the teacher's own teaching. It is important for teachers to reflect on their own teaching to improve on it. To make such a reflection, the teachers are expected to ascertain the degree of students' understanding of what has been taught. Therefore, a correlation is suggested between the two processes.

Creating a comfortable environment for teachers' test scoring reduces the workload and makes ascertaining the degree of students' understanding easier. The system proposed by Nakagawa et al [17] reduces the work time in all scoring processes for elementary schools and junior and senior high schools. By using this system, the work processes of scoring, calculation, and scoring records are automated and are expected to significantly reduce the

teacher's burden. Meanwhile, several researchers reported difficulties among teachers to ascertain students' level of comprehension through the implementation of a scoring support system [20][21]. Hence, Nakagawa et al. [20] introduced a method in which teachers check the scored answer sheets after using the system to determine students' level of knowledge, while Nakagawa et al. [21] considered adding a preview function for answers to the system. This points to the need for future studies to explore this scoring support system alongside the impact of using such a system to evaluate students' degree of understanding and reflect on one's own teaching. In the future, it is expected that computer systems will be applied to various tasks of teachers to resolve the busyness of teachers. However, we need to be mindful that the introduction of computers in school work would not only improve efficiency in terms of work but should also assist and improve teachers' teaching. This outcome will depend on how we proceed with the future digital transformation in education.

Acknowledgement

We would like to express our appreciation to teachers for responding to the survey. We are also particularly grateful to Yagisawa, F., Murai, A., and Asato, M., for their encouragement and continuous support for the study.

References

- [1] Ministry of Education, Culture, Sports, Science and Technology, "Shogaikoku no Kyōiku Tōkei 2019 Nenban", https://www.mext.go.jp/b_menu/toukei/data/syogai-koku/_icsFiles/afieldfile/2019/08/30/1415074_0_1.pdf, 2019. (accessed March 23, 2022)
- [2] Ministry of Education, Culture, Sports, Science and Technology, "Heisei 30 Nendo Gakkō Kihon Chōsa (Kakuteichi) no Kōhyō ni Tsuite", https://www.mext.go.jp/component/b_menu/other/_icsFiles/afieldfile/2018/12/25/1407449_1.pdf, 2018. (accessed March 23, 2022)
- [3] Ministry of Education, Culture, Sports, Science and Technology, "Gakkyū Hensei oyobi Kyōshokuinteisū ni kansuru Shiryo 2005", https://www.mext.go.jp/b_menu/shingi/chousa/shotou/029/shiryo/05061001/sankou002.pdf, 2005. (accessed March 23, 2022)
- [4] Research Institute for School Education, "Shogaikoku no Kyōiku no Jōkyō", Gakko Tosho Co., Ltd., Tokyo, 2006.
- [5] Ministry of Education, Culture, Sports, Science and Technology, "2007 White Paper on Education, Culture, Sports, Science and Technology", https://warp.ndl.go.jp/info:ndljp/pid/11293659/www.mext.go.jp/b_menu/hakusho/html/hpac200701/1283225_001.pdf, 2008. (accessed March 23, 2022)
- [6] Stevenson, H. W. & Stigler, J. W. "The Learning Gap: Why our schools are failing and what can we learn from Japanese and Chinese Education", Simon & Schuster, New York, 1992.

- [7] Andreas Schleicher, "PISA 2018 Insights and Interpretations", <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>, 2018. (accessed March 23, 2022)
- [8] National Institute for Educational Policy Research in Japan. "The Work Environment of Teachers in Japan", <https://www.nier.go.jp/English/educationjapan/pdf/20180816WEJT.pdf>, 2018. (accessed March 23, 2022)
- [9] A. Bannai, S. Ukawa and A. Tamakoshi, "Long working hours and psychological distress among school teachers in Japan", 2015, Vol. 57, pp.20
- [10] M. Suzuki, "Teachers' health and labor reform", JJHEP, 2018, Vol. 26, No. 3, pp.298-304.
- [11] T. Okumura, T. Miyashita, A. Masui, K. Mori, M. Kitajima, "Shinnin-Kyōin niokeru Burnout Keikō to Rishoku-ishi no Keijiteki-henka", The 61st Annual Meeting of Japanese Association of Educational Psychology, 2019, p.390.
- [12] Ministry of Education, Culture, Sports, Science and Technology, "Kyōin Kinmu Jittai Chōsa (Heisei 28 Nendo) (Kakuteihi) ni tsuite", https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/079/siryo/_icsFiles/afieldfile/2018/09/28/1409717_4_1.pdf, 2017. (accessed March 23, 2022)
- [13] Libertas Consulting Co., Ltd., "Kōritsu Shōgakkō & Chūgakkō Kyōin Kinmu Jittai Chōsa Kenkyū Hōkokusyo, Report on Survey Contracted from Ministry of Education, Culture, Sports, Science and Technology", http://www.mext.go.jp/component/a_menu/education/detail/_icsFiles/afieldfile/2018/09/27/1409224_005_1.pdf, 2018. (accessed March 23, 2022)
- [14] Benesse Educational Research and Development Institute, "Dai-5-kai Gakushū Shidō Kihon Chōsa (Shōgakkō/Chūgakkō-ban) [2010]", <https://berd.benesse.jp/shotouchutou/research/detail1.php?id=3243>, 2010. (accessed March 23, 2022)
- [15] T. Takeuchi, A. Sakuma, "Development of Marking-support Soft in Task-given-type Examination", Journal of Japanese Society for Information and Systems in Education, Vol. 21(1), 2004, pp.33-42.
- [16] X. Liu, Y. Aoki, "Kyōtsu Kamoku Kyōiku ni okeru Saiten System Kaihatsu oyobi Jissen," Computer & Education, 2004, Vol. 16, pp.38-42.
- [17] S. Nakagawa, K. Sato, R. Saito, T. Horita, "Deployment and Evaluation of Scoring Support System in Written Test for Subject Teachers and Classroom Teachers", Japan Society for Educational Technology, 2020 Vol.43, No.4, pp.433-445.
- [18] Ministry of Education, Culture, Sports, Science and Technology, "Case Studies of School Work-Style Reform in Japan Revised Edition", https://www.mext.go.jp/content/20220221-mxt_kouhou01-000020595_1.pdf, 2022. (accessed March 23, 2022)
- [19] E. Kajita, "Keiseiteki na Hyōka no tameni", Meijitoshō Shuppan Corporation, 1986.

- [20] S. Nakagawa, S. Itagaki, T. Horita, “A Preliminary Study of the Use of Scoring Support System in Writing Test on Teacher’s Grasp of Learners’ Understanding and Instruction in Junior High School”, Research Report of JSET Conferences, 2020, JSET20-4, pp.191-198
- [21] S. Nakagawa, R. Saito, S. Itagaki, K. Okubo, M. Osaki, N. Yamazaki, T. Horita, “A Study of Development of Scoring Support System’s Function to Grasp Understanding Status of Individual Learners in Primary and Secondary Education”, JSiSE Research Report, 2022, Vol.36, No.6(2022-3), pp.167-170
- [22] Systems Research & Development Institute of Japan, “Gakushū Shidō to Gakushū Hyōka ni taisuru Ishiki Chōsa Hōkokusho. Report on Survey Contracted from Ministry of Education, Culture, Sports, Science and Technology, 2009”, http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/siryu/_icsFiles/afieldfile/2010/02/19/1289879_1.pdf, 2010. (accessed March 23, 2022)
- [23] Benesse Educational Research and Development Institute, “Dai-3-kai Gakushū Shidō Kihon Chōsa”, <https://berd.benesse.jp/shotouchutou/research/detail1.php?id=3249>, 2002. (accessed March 23, 2022)
- [24] Benesse Educational Research and Development Institute, “Dai-5-kai Gakushū Shidō Kihon Chōsa (Kōkō-ban) [2010]”, <https://berd.benesse.jp/shotouchutou/research/detail1.php?id=3245>, 2010. (accessed March 23, 2022)
- [25] The Japan Association for Research on Testing, “Minaosō, Test wo Sasaeru Kihon no Gijutsu to Kyōiku”, Kanekoshobo, Tokyo, 2010.
- [26] Gakkō Kyōzai Katsuyō Shidōhō Kenkyukai, “Veteran Sensei Jikiden Work Test no Katsuyōhō,” Kyoikudojinsha, Tokyo, 2014.
- [27] S. Nakagawa, T. Horita, “A Study on Analysis of Scoring Work and Thinking during Scoring for Subject Teachers in Junior high school”, Research Report of JSET Conferences, 2019, JSET19-4, pp.239-244