Significance and Challenges of Formative Ipsative Assessment in Inquiry Learning: A Case Study of Writing Activities in a “Contemporary Society” Course in a Japanese High School

Kohei Nishizuka

Abstract
The theory of formative assessment highlights the importance of ipsative assessment, which uses internally generated criteria and goals to compare students’ performance with their past performance. This study designed a formative ipsative assessment practice as part of inquiry learning in high school and clarified its significance and challenges. The results of three rounds of academic writing for high school freshmen showed that the practice gradually improved students’ overall writing skills; thus, qualitative improvement and transformation in performance-centered learning were expected. In high schools, where a single teacher is responsible for many students and is constrained by external standards such as university entrance exams, one strategy is to align learning and assessment by developing abstract higher-order criteria and then developing them into concrete lower-order criteria appropriate for each student. This paper presents five challenges and these solutions.

Keywords
ipsative assessment, formative assessment, criteria, inquiry learning, academic writing

Social Background
The process by which teachers and students become aware of the progress and needs of teaching and learning and use the assessment to improve the quality of their educational activities is called formative assessment (Note 1). Until now, the theoretical framework of formative assessment in Japan and abroad has been based on a context in which teachers present students with standards (e.g., goals as socially regulated values) and criteria (e.g., tools for judgment, qualitative characteristics of learning objects) and guide them toward achieving predetermined goals (James, 2017; Nishizuka, 2020; Sadler, 1987, 2007). Criterion-referenced assessments are expected to have high cost-effectiveness and effect sizes in terms of test score gains (Black & Wiliam, 1998; Hattie, 2009).

However, several pitfalls of this concept have been observed. First, its purpose is to support learning that is only valid for schooling and to strengthen and reproduce the knowledge base of existing communities (Engeström, 2015). The socio-cultural aspects of assessment have been widely discussed (e.g., Elwood, 2006; Torrance, 2012), within which it has been argued that while learning improvements are aimed at developing autonomous, critical learners, in reality, assessment has been strongly influenced by behaviorist views of learning. That is to say, the assessment procedures make students passive and obsessed with passing exams and winning competitions. Formative assessment—with its primary focus on winning a competitive testing system—has stifled creative and innovative learning and narrowed the meaning of learning. Second, teachers and students lack understanding of assessment criteria and conflate age and developmental stage as coinciding (Fleer, 2015). Sharing criteria with students has been considered effective in improving work practically (Sadler, 1989). However, because teachers can develop criteria based on learning objectives and standards, “sharing” has been replaced by “transmitting” from teachers to students (Christodoulou, 2017). Assessments cannot improve learning if students are neither familiar with the criteria nor motivated by them.
Thus, unattainable goals and incomprehensible criteria for students make assessment unwieldy, which narrows the definition of learning. To warn about this chaos, Hughes argued that assessment should contribute toward enriching students’ active motivation and learning of how to learn toward becoming lifelong learners. To that end, learning goals must be internal (not criterion-based) and assessment must be linked to the starting point of learning, without comparison with others (not norm-based) (Hughes, 2014, p. 73), and for the need for assessment that uses students’ internally generated criteria and goals to continually build comparisons with their past performance. This concept is called ipsative assessment, which applies the idea of self-referential assessment in psychometrics to the level of classroom practice.

Although ipsative assessment only has a short research history, literature review suggests that it has three advantages, primarily because of the alternative assessment reforms in higher education. First, ipsative assessment supports the development of complex higher-order cognitive functions, such as academic writing and evaluative judgments, which cannot be measured using traditional memory-replay tests (Malecka & Boud, 2021; Malecka et al., 2021; Seery et al., 2019; Univio & Pérez, 2019). Second, as the school systems are part of the global neoliberal marketplace, which intensifies the competition between students and between schools, an ipsative perspective provides a socially safe space for people to respect their own identities and develop positive, social emotions, such as resilience and curiosity (Lucas et al., 2013; Rattray, 2018). Third, past and present learning comparisons that are based on various learning resources are possible, such as e-portfolios (Malecka et al., 2021) and self-administered testing in virtual learning environments (Penn & Wells, 2018), and extensions to online learning environments are encouraged, which adds value to the assessments. Therefore, unlike criterion-referenced assessments, which involve authoritative and political control, ipsative assessments are less reliant on teacher-imposed task performances and promote rounded learning, rather than focusing on learning targets only legitimated by summative tests, which allows students to be more certain about their futures and to better understand the value of their learning.

Formative feedback, the core of formative assessment, should provide information on educational improvement, which Hattie and Timperley classified into three categories (Hattie & Timperley, 2007): feedup, answering “where do we need to go”; feedback (FB), answering “where are we now”; and feedforward (FF), answering “where do we need to go next.” In other words, FB is about inferring the distance between past and present developmental levels, and FF is about working to close the gap between the present developmental level and the future developmental level defined by the feedup. In formative ipsative assessment, in principle, FB and FF are repeated for the same learning task to gradually improve the quality of learning. The differences between formative ipsative and criterion-referenced assessment are summarized in Table 1.

### Table 1. Characteristics of Formative Ipsative and Criterion-Referenced Assessment (Hughes, 2014; Nishizuka & Arimoto, 2019).

| Formative assessment | Criterion-referenced | Ipsative (Self-referential) |
|----------------------|----------------------|-----------------------------|
| **1. Purpose and goal setting process** | | |
| Purpose | Achievement of standards | Maximization of personal competence and motivation |
| Meaning of goal | Standards for accountability | Set up and transform so that individually optimized goals are connected to learning |
| **Main target** | | |
| Group | | Individual |
| **Origin of zone of proximal development (ZPD)** | Zone between current performance and external goals | Zone between past performance and current performance |
| **Definition of development** | Process for achieving goals | Process for transforming self (achieving goals) |
| **2. Criteria: tools used in assessment** | | |
| Design | Pre-set irrespective of students | Generate criteria based on student development |
| Clarification | Often explicitly stated | Not always explicitly stated |
| Developmental scale | Scales are common to all students and basically unchanging | Scales vary by individual and are diverse |
| **3. Compatibility in implementation** | | |
| Summative aspect | Effectiveness reflected in (standardized) test scores | Need moderation and standardization |
| Socioemotional aspect | Often limited to measuring cognitive abilities | Inclusion of emotions to control cognitive abilities |
| Preferred school stage | Mainly secondary education | Mainly primary education and higher education |

Research Question

Formative ipsative assessment is characterized by referring to various criteria and encouraging individually optimal information for improvement. For this reason, elementary schools, universities, and free schools, where learning tends
to be constrained by entrance exams and other high-stakes tests, are considered desirable practice environments (Nishizuka & Arimoto, 2019). Thus, how can ipsative assessment be practiced in the context of secondary education? In this paper, the author focused on inquiry learning.

According to Kirschner et al. (2006, p. 76), inquiry learning has two characteristics: students “solve ‘authentic’ problems or acquire complex knowledge in information-rich settings based on the assumption that having learners construct their own solutions leads to the most effective learning experience,” and “knowledge can best be acquired through experience based on the procedures of the discipline.” Pedaste et al. (2015) conducted a literature review of inquiry learning procedures. They observed that inquiry learning involves four steps: orientation (attracting students’ interest in social issues), conceptualization (asking questions and generating hypotheses), investigation (examining, testing, and discussing), conclusion (because as communication and reflection are involved in all four processes). Bianchi and Bell (2008, p. 27) categorized inquiry learning in terms of teacher intervention into four stages: confirmation inquiry (students confirm principles through activities in situations for which outcomes are known); structured inquiry (students follow pre-defined procedures to investigate questions the teacher posed); guided inquiry (students follow self-designed and selected procedures to investigate questions the teacher posed); and open inquiry (students follow self-designed and selected procedures to investigate questions the student posed). Thus, inquiry learning is characterized by creating a learning space between the student and social context, using knowledge and abilities through performance, continuously reflecting on thoughts and actions and creating unknown knowledge that becomes a new tool for the student to solve problems.

The support of inquiry learning is commonly referred to as guidance, the importance of which has been pointed out. Guided inquiry learning is considered the most effective learning strategy. Criterion-referenced tests have shown that guided inquiry learning contributed to improved scores, and guided and open inquiry learning was more effective than knowledge-infused lecture-based teaching (Conway, 2014; Furtak et al., 2012; Margunayasa et al., 2019); in addition, guided learning is more effective for open inquiry learning (Alfieri et al., 2011; Chen & Yang, 2019; Furtak et al., 2012; Kwietniewski, 2017; Lazonder & Harmsen, 2016). Alfieri et al. (2011) identified three key points for inquiry task creation: tasks must provide scaffolding support, must allow students to explain and check the accuracy of their thinking through timely feedback, and must provide an example for successful learning. Lazonder and Harmsen (2016) organized inquiry guidance types into process constraints that restrict the comprehensiveness of the learning task, status overviews that make the task progress and/or the learning visible, prompts that act as reminders to perform an action, heuristics that serve as reminders to perform an action and suggestions on how to perform that action, scaffolds that explain or take over the more demanding parts an action, and explanations that specify exactly how to perform an action. Kwietniewski (2017) also emphasized that project-based learning instructions required differentiation to ensure individual optimal coordination. Therefore, based on these various guidance methods, the role of formative assessment in inquiry learning can be summarized as follows (Barron & Darling-Hammond, 2010; Grob et al., 2017; Harlen, 2003): defining the task and articulating the learning movement toward a criterion; diagnosing a student’s current developmental level to accurately identify the learning variety; providing formative feedback to each student to encourage the next learning step; and training students to feel confident in conducting self-regulated learning. Given these requirements, the provision of continuous support is a prerequisite for effective inquiry learning.

However, criterion-referenced assessment of formative function remains problematic as described in Social Background. In one view (James et al., 2007), the incompatibility of inquiry learning with predetermined goals and criteria was suggested. Particularly, overly rigid criteria for knowledge-creating learning can substantially narrow the view of learning needs and feedback, where there are problems of teachers’ fear of losing control over student learning and students’ inability to advance beyond the learning envisioned by the teacher because teachers are preoccupied with achieving standards that prove the quality of learning (Engeström, 2015). Preventing such suppression of learning by assessment requires efforts to make assessment criteria more aligned with student learning (Nishizuka & Arimoto, 2019; Torrance, 2012).

As previous research has highlighted that guidance for students and teachers is vital to inquiry learning and that ipsative approaches could be suitable toward solving inquiry learning assessment problems, it could be hypothesized that ipsative approaches are effective and have educational value. Therefore, this study attempted to design ipsative assessment practices as part of inquiry learning in a Japanese high school and concretely elucidate their significance and challenges by analyzing the learning effects and their influence factors.

**Method**

**Case Study**

The data for the case study were collected from the civics class “Contemporary Society” of Teacher Aizawa (pseudonym), who worked at one private 6-year secondary school in Japan (hereafter, the School) in the 2019 academic year (Note 2). He taught “Contemporary Society” for 154 students in four of the five classes of first-year high school students; Teacher Urakawa (pseudonym) taught the remaining class. The two teachers held weekly meetings to share instructional content and ipsative assessment practices and
make adjustments to align the classes. The yearlong curriculum is designed to be flexible and combines classroom lectures and discussions (Table 2). This class at the School meets on average twice a week (50 minutes for each class) and is required for all first-year high school students. In the first class, Teacher Aizawa introduced the purpose of the class to the students. The presentation overviewed how social structures, such as, Japan’s declining economic power and birthrate, aging society, and multicultural society will change when the students enter the workforce.

Ipsative assessment must be integrated into a systematic curriculum and student learning activities. Over the year, the students learned about the positives and negatives of global capitalism, such as the widening gap between the rich and poor, and the preponderance of child labor because of income and economic inequality. Therefore, Teacher Aizawa and the author created the following report assignment regarding social disparity. We ipsatively tracked students’ development during the first semester to the middle of the second semester of the 2019 school year. The report writing also served as the basis for a problem-solving course in which students had to pose their questions in the third semester.

In the first half of the first semester, students explored disparity in global capitalism. They played the trading game and read *Capital in the Twenty-First Century* by Thomas Piketty, deepening their understanding of the definition and causes of disparity and the social structure that accelerates disparity. Toward the middle of the lecture, Teacher Aizawa modified a university entrance exam question and assigned its original task to the students. It asked the students to state their position on disparity and to criticize the question asked in the examination. In the conference with other social studies teachers on October 9, 2019, he explained his intentions.

The question in the essay was to present a solution to the growing disparity, but I thought that if I just answered that question, it would be a disciplined writing in line with the aim of the question. I asked them to write a report on what they think about the disparity, after analyzing the assumptions, values, or stance of the university that is asking them to present a solution.

The task was as follows.

This is the question and materials of a university’s second-semester essay examination. Analyze what assumptions you think question no.2 is based on. Additionally, clarify what you think about this university’s position and discuss the disparity.

Question no.2 stated, “describe what prescriptions you can draw for the widening disparity in the world, and give your thoughts on them in 700 to 800 words.” Four materials were presented: a newspaper article explaining Piketty’s book, an article critical of Piketty’s argument for a wealth tax, a section of a book on global tax trends, and four graphs from Oxfam that provided evidence of the growing scale of disparity.

Table 2. “Contemporary Society” Curriculum in FY2019 (Until the Middle of the Second Semester).

| No. | Theme | Contents |
|-----|-------|----------|
| First semester of FY 2019 (April–July) |       |          |
| 1   | Introduction | Introduction presentation, and activities to consider the underlying issues |
| 2–3 | Trading game (1) (2) | Trading game (Note 3) |
| 4   | Trading game (3) | Reflection of the trading game (understanding the disparity between nations) |
| 5   | Trading game (4) | Disciplines of global capitalism |
| 6–8 | Trading game (5) (6) (7) | Sharing disciplines and discussion, and understanding background knowledge of the disciplines. |
| 9   | Trading game (8) | Examining the first half semester |
| 10  | $r \geq g$ (1) (Note 4) | Analysis of dominant disparity (*Thomas Piketty Capital in the Twenty-First Century*) |
| 11  | $r \geq g$ (2) | Exploration of data |
| 12–13 | $r \geq g$ (3) (4) | Watching a video on poor high school girl and discussing the rectification of social disparities |
| 14–15 | $r \geq g$ (5) (6) | Reading the materials on Japanese relative poverty |
| 16–18 | Report (1) (2) (3) | Reading essays from the university entrance exam, writing and submitting the first report |
| 19  | Comment on the report | Explaining the criteria for the report, instructions for rewriting the first report |
| Second semester of FY2019 (August–December) |       |          |
| 20–21 | Mock election (1) (2) | Analysis of political parties |
| 22   | Mock election (3) | Analysis of political parties’ issues |
| 23   | Mock election (4) | Review of the electoral system |
| 24   | Mock election (5) | Discussion point review and identification of candidates |
| 25–27 | Report (4) (5) (6) | Returning comments on second the report, creating a thesis structure and gathering arguments |
| 28–31 | Report (7) (8) (9) (10) | Writing and submitting the third report |
| 32   | Mock election (6) | Data analysis of mock election results |
| (omitted) |       |          |
**Experimental Design**

Following standard practice for prior research to adopt Hughes’ theoretical framework, this study’s experimental design used a basic scheme proposed by Hughes (2014): the iterative process of ipsative feedforward (IFF) and ipsative feedback (IFB) (Figure 1). The submission period of each report was approximately 90 days. The submission and return dates for the IFF and IFB reports were early June and mid-June, late July and early September, and late October and mid-December. The students handwrote the first draft and second draft reports on the left and right sides of the paper, respectively. On the back of the paper, the students reported what and how they improved from the first draft to the second draft (SIFB1). The third draft report was an online submission with text input.

First, the teacher needed to provide a comparison paper for the students to improve their first draft, because no prior learning products were available. In addition, assessing approximately 150 students was too difficult for one teacher.

**Figure 1.** Cumulative scheme of ipsative assessment (based on Hughes, 2014, p. 82).

**Table 3.** Rubric based on the first draft (showing level 3).

| Major item | Description | Minor item | Level 3 (outstanding) |
|------------|-------------|------------|-----------------------|
| A. critical reading comprehension | Are able to understand the opponent’s argument and its basis and then consider its premise, underlying values, and arguments and examine its validity. | A1. accurate reading of materials and questions | Are able to accurately understand not only individual arguments in materials and discussions but also their rationale, assumptions, and relevance. |
| | | A2. critical reading of materials and questions | Are able to examine the validity of not only the individual claims in the materials and discussions but also the arguments, premises, and relevance of those claims. |
| B. logical writing | Clearly articulates one’s own claims and arguments, and logically connects the claims with each argument and specific examples. How the argument is formulated accurately captures the complexity of social phenomena. | B1. definition and reasoning | Are persuasive because they clearly define the substance of the concepts they describe and base their arguments on highly generalized arguments. |
| | | B2. development of logic | Are consistent in the development of logic and the ability to carefully investigate complex objects and explain them. |
| C. integrated knowledge management | Are able to think in a way that connects the studies of the subject with studies of other subjects, research on issues, and current knowledge. | | In addition to the perspectives and knowledge gained in the past “contemporary society” classes, students can also use information gained from other subjects, research projects, and the news in the discussion. |
| D. writing style | Are writing based on the basic rules of writing. | | The writing is clear, with no twists or turns in the main idea, and is easy to read, with paragraphs divided by main ideas and a one-letter space at the beginning of each sentence. Appropriate vocabulary and expressions for logical narration are used. |
Therefore, Teacher Aizawa and the author created a three-level rubric based on the contents of the first draft (Table 3). Teacher Aizawa returned the first draft without commenting on it. Subsequently, during the 19th class, called “Comment on the report” (TIFF1), he explained each criterion of the rubric, provided several examples, and asked the students to submit the improved version.

When the second draft was submitted, the students described the improvements they made (SIFB1). Subsequently, during the students’ summer vacation period (late July to mid-August), Teacher Aizawa prepared detailed FB (TIFB1) and FF (TIFF2) for each student in text format and distributed them to the students after the vacation. All students wrote the third draft in approximately 1 month, receiving several writing periods during class. The bottom of the students’ third draft provided space to describe their improvements and compare the paper with the second draft (SIFB2), free writing space to reflect the transformations and achievements they felt through the three drafts (SREF), and negative writing space to report what they had difficulty with or did not understand (SNEG) (Table 4).

This study also included a written interview with Teacher Aizawa conducted after the experiment (January 4, 2020), an oral group interview with Teacher Aizawa and Teacher Urakawa (July 4, 2019), and recorded data from a conference (October 11, 2019) in which full-time social studies teachers reported on their teaching and progress. Thus, the value judgment of this practice from the teacher’s viewpoint was included in the analysis.

### Analytical Method

Although the experimental design was based on Hughes (2014), this study was characterized by a specific analytical method. A majority of the methods used in previous and related studies (Hughes, 2014, 2017) either quantitatively analyzed the data or interpreted qualitative data based on the arbitrariness of the analyst. However, although the quantitative analyses were based on an ipsative transformation of the numeric forms from self-report questionnaires, the performance quality reflected in the numeric data was limited because of the lack of the elicitation of the skills gained from the learning process. In addition, while it was easier to ascertain student transformations from qualitative data because of the small number of subjects as most research fields were in higher education, there were few ipsative assessment practice examples for larger student and teacher populations. Therefore, to more convincingly examine the effects of, and associated factors for, ipsative assessment, it was important to avoid arbitrary data processing and interpretations and focus directly on students’ skills and the content of teacher–student feedback and feedforward. With these issues in mind, this study employed a mixed research approach, wherein the qualitative data were analyzed both quantitatively and qualitatively, which allowed for the learning work, process, feedback and feedforward, and related teachers’ school activities to be the direct analysis targets.

The analysis was divided into three: the students’ learning outcomes, the factors behind the outcomes, and the challenging aspects of the ipsative assessment.

### Students’ learning outcomes

Three methods were mixed to measure students’ learning outcomes. To objectively analyze student reports, the author used the KH Coder (3. Beta. 03a) to conduct a quantitative analysis of qualitative data by multivariate analysis (Higuchi, 2020). All handwritten first drafts and second drafts were transcribed into text by the author, and one draft was copied into one cell in Excel (Microsoft Office). As part of the cleansing (pre-processing) of the report, orthographical variants of words in kanji and hiragana were corrected (i.e., varied Japanese iterations similar in meaning were considered as one), and numeric symbols and question marks that could not be processed by KH Coder were replaced with simple numbers and punctuation marks (Higuchi, 2020). Figures and tables inserted in each report were excluded from the analysis. Correspondence analysis was conducted with each report set as an external variable. In the analysis, extracted words are sorted so that the correlation between rows and columns takes the maximum value, and the relationship between extracted words in each row and column is shown in a scatter diagram. It is frequently used for visually exploring commonalities and differences between reports because it simultaneously distributes relationships between words found in various data.
The extracted words were mapped into four quadrants and two dimensions (row and column), and the characteristics of these two dimensions were interpreted and named on the basis of the words and text. The analysis conditions were determined based on the conditions that could be most objectively and validly interpreted from the relationships between the textual content and the diagrams that had been mapped in the correspondence analysis.

Next, the reflections of students and teachers are examined. The students were asked to reflect on their learning transformation by writing three reports at the time of submission of the third draft (SREF) by using KH Coder’s co-occurrence network for the analysis. Because of the large amount of text in the SREF, there was a risk of arbitrary interpretation in manual after-coding; thus, the co-occurrence network that visualizes the relationship between words in the text by using multivariate analysis was selected. As with correspondence analysis, detail conditions depend on analysts’ judgment and in turn are based on valuable and interpretable relations of text data and results of analysis.

Last, qualitative data regarding students’ transformation, which were acquired by a written interview with Teacher Aizawa conducted after the experiment, were used to combine the two aforementioned analytical results.

**Factors behind the outcomes.** Three methods were adopted to explore some of the factors leading to the learning outcomes. Teacher Aizawa and the author extracted the problems for improvement that many students had and created a three-level (3: outstanding, 2: satisfactory, and 1: needs improvement) rubric with higher-order criteria to facilitate a wider learning range (Table 3). Notably, this practice is an incomplete ipsative assessment because the rubric was translated from ipsative goals and criteria provided directly to individual students for the goals and criteria at the group level.

To analyze the formative feedback from the teacher, the feedback analysis tool (FAT) created by Hughes et al. (2015) was used. FAT comprises 10 codes (Table 5). First, praise and recognizing progress (IFB) are distinguished by whether they are rewards for results or for progress. Critique comprises pointing out errors when checked against existing official rules, such as those for grammar and spelling, factual critiques of content, and critiques of the approach to structure and argument. Advice (IFF) was divided into three categories: specific, content-centered suggestions for the task students are working on; general points that refer to the current assignment; and advice for future tasks. Clarification requests or queries are defined as asking for definitions of words or checking facts; those that are not in these categories are considered unclassified statements. For example, the author excerpted the relevant descriptions from the material (in TIFF1) that Teacher Aizawa explained in the 19th class while showing the rubric.

The teacher’s formative feedback (TIFB1 and TIFF2) on the second draft reports was organized on the basis of the
rubric and FAT. Teacher Aizawa judged each formative feedback comment (including duplicates), determined the appropriate rubric criterion, and then communicated this to the students. After the experiment was completed, the author reclassified each comment into FAT.

The focus then shifted to the SIFB content to investigate the TIFB and TIFF student contributions. Students were asked to compare the transformations between the first and second drafts (SIFB1) and the second and third drafts (SIFB2), respectively. After the individual comments were inductively categorized by content to create categories, they were applied to each criterion in the rubric. The actual numbers and percentages were then calculated.

In this practice, because the second draft reflected the self-assessments and the third draft reflected the teacher’s assessments, it was expected that the assessors, that is, the teachers and students, would influence the learning. The student improvement characteristics could also be analyzed on the basis of the extent of the revisions they had made to their reports. Therefore, the author confirmed the change in writing content from the first draft to the second draft and from the second draft to the third draft by using the percentage of sentence concordance. For cleaned reports submitted on time from the first draft to the third draft, the concordance rate (copy-and-paste ratio) of expressions was verified between the first and second drafts (1→2 group) and between the second and third drafts (2→3 group), using the copy-and-paste judgment support software “Copypamer V5” (Ank Co., Ltd.). The copy-and-paste rate is calculated by dividing the number of suspected copy-and-paste clauses by the total number of clauses in the document. The 1 clause is a clause separated by punctuation, reading marks, commas, periods, line breaks, and parentheses.

Challenging aspects of ipsative assessment. The interviews with Teacher Aizawa and Teacher Urakawa and the students’ free descriptions of the SNEG in the third draft report were included for this analysis to assess the challenges and propose possible solutions. These aspects were then inductively categorized to explore as many of the ipsative assessment issues as possible.

Results and Discussion

Learning Outcomes

This section analyzes the learning outcomes from the student reports. The analysis was based on 130 first drafts, 130 second drafts, and 141 third drafts submitted by the deadline. In correspondence analysis, the author automatically extracted the top 70 words with significant differences. The minimum number of occurrences is 185, as all words being mapped had a cumulative total of 185 and more uses. The minimum number of documents is 1, which makes the word to be analyzed because it appears in one or more drafts. The unit of analysis is H5 (a cell in the Excel sheet), which was expected to have the most combinations of co-occurrence relations. In addition, to remove words with few characteristics concentrated near the origin (0, 0), only the top 70 words farthest from the origin were plotted (−3.04 ≤ Dimension 1 ≤ 3.57, −2.17 ≤ Dimension 2 ≤ 2.80). As there were still words extremely far from the origin, the overall balance in the figure was automatically accounted for, which revealed many extracted words still clustered near the origin. Therefore, the “zoom around the origin” function in KH Coder was applied to rearrange the extracted words densely clustered near the origin so that they could be more easily analyzed without changing the analysis results (Figure 2). The “zoom around the origin” function deleted the coordinate information for each word in the analysis, with the closer the extracted words to each other the stronger the relationship and the less distinctively scattered the words at the origin.

As mentioned before, each Dimension is named by the analyst. In Dimension 1, “think (思う),” “rich (豊か),” and “money (お金)” are characteristic above zero (upper half of Figure 2). On the other hand, “think (考え,る),” “understand (分かる),” “increase (増える),” “many (多い),” “big (大きい),” and “wealthy (裕福)” are characteristic below zero (lower half of Figure 2). For example, I think (思う) it is okay to have disparities (格差). However, I do not think (思う) it is good (良い) for people (人々) who do not have money (お金) to live on some cents per meal or be unable to go to high school or university (大学) because of lack of money (お金). If we could raise the necessary minimum for life (生活) that we have now (今), I do not think (思う) it would matter (問題) how much the disparity (格差) widens (広がる).

(first draft, Student 51)

I think (考え,る) there is a way to eliminate disparity (格差) and make everyone equal (平等), but I think (考える) that disparity (格差) is good (良い). [...] However, no matter how much disparity (格差) there should be, it is not good (良い) to have people (人々) living on some cents per meal or not having a house. If the lowest income (収入) line is raised, people will get what I consider a normal life (生活). I think (考える) it is important to raise the overall income (収入) of each individual first. (second draft, Student 51)

From a study published by the United Nations in 2012, a global tax (グローバル・タックス) would give us the money we need (必要) to solve global problems on an annual basis. Additionally, a comparison of the reduction in greenhouse gas emissions between countries with and without a carbon tax (税) clearly shows (分かる) that countries with a carbon tax (税) have reduced their emissions by a greater (多い) percentage. (second draft, Student 70)

The first draft was characterized by the use of writing styles and expression techniques not appropriate for a thesis statement, and the second draft used descriptions that fit the format of an essay to analyze what can be understood from the figures and tables. Therefore, an index of subjective
explanation–objective explanation was assumed for Dimension 1 (Figure 2).

Below zero on Dimension 2 (left half of Figure 2) are, for example, “effort (努力),” “equality (平等),” “questions (設問),” “materials (資料),” “figures (図),” and “Piketty (ピケティ).”

In addition, the disparity (格差) is not on a scale that we can do anything about right now, but I thought (思う) with a little effort (努力), we can change it. (first draft, Student 9)

I agree with Piketty’s (ピケティ) ideas because I believe (考える) that the best way to reduce global (世界的) inequality is to reduce the wealth of the rich (富裕層) and increase (増える) the wealth (富) of the poor (貧しい). (second draft, Student 13)

The words above zero on Dimension 2 (right half of Figure 2) comprise “poverty (貧しい),” “wealth (富裕),” “education (教育),” “income (所得),” “taxation (課税),” “livelihood (生活),” “policy (政策),” “economy (経済),” and “distribution (分配).”

The bad (良くない) thing about disparity (格差) is that there are people (人) who have so little money (お金) that they cannot afford to live (生活). The relative poverty (貧困) rate in Japan (日本) is increasing (増えている) every year (毎年). [...] I hope (考える) this relative poverty (貧困) rate can be reduced. To that end, we need (必要だ) to raise the minimum wage and redistribute (再分配) money. I think (考える) it is important not only to raise the minimum wage but also to raise it such that the difference (差) between the two is eliminated and that all places in Japan (日本) have the same minimum wage. Also, [...] at the graph (グラフ) “Trends in Regional Disparities (格差) in Minimum Wages,” I can see (分かる) that the difference (差) between the highest and lowest minimum wages is getting larger (大きくなっている) every year (毎年). For these two reasons (理由), I believe (考える) it is not necessary (必要) to eliminate economic disparity (経済格差), but it is important for everyone to have a certain amount of money through redistribution (再分配) and raising the minimum wage. (third draft, Student 51)
their claims. In the third draft, a wide perspective for analysis and precise analysis of the disparity solution was evident. The students were mainly attempting to analyze the current state of education and policy, demonstrate the issues, and logically explain the proposals derived from the analysis. Therefore, the components on Dimension 2 can be named as an index of solution proposal with little evidence and precise analysis and solution proposal (Figure 2).

Therefore, the first draft used many subjective and poor expressions and irresponsible proposals for solving the disparity. In the second draft, many objective expressions, opinions, and rationales, for the materials and questions were stated, but there was no significant change in the content (this point is further analyzed in the next subsection), which was limited to thin rationales and superficial proposals for solutions. The third draft contained a good balance of subjective and objective explanations; a wide perspective for analysis and precise analysis of disparity resolution was thus evident. The students were mainly attempting to explain their analysis of the current situation of education and policy, pointing out the issues and the suggestions that could be drawn from it.

Next, the reflections of students and teachers are examined. Figure 3 shows the co-occurrence network of the SREF’s descriptions of 124 students who submitted their work on time. The minimum number of occurrences was set to 5, the minimum number of documents was set to 1, and the unit of analysis was set to H5; frequent words were automatically extracted. The Jaccard coefficient was used to measure the similarity of the texts, and only the top 50 words were extracted. The KH Coder has two possible word extraction methods: extract the words in the order of increasing Jaccard coefficient, and only extract words with Jaccard coefficients higher than a specified value. In this analysis, the first method was adopted. As the quantity of data to be analyzed was relatively small and the SREF was diverse, the results could not be fully interpreted by only examining the extracted words that had high Jaccard coefficients.

The numbers on the lines between the extracted words are the coefficients of co-occurrence relations. When classifying the extracted words into categories (community structure of the network), the method called modularity resulted in 02, 03, and
Category 04 in Figure 3 being combined into one category. It was clear that there were clearly different kinds of categories mixed within this one category. Therefore, the author used a method called random walks (Azizifard, 2014), which clearly separated the categories of 02, 03, and 04 in Figure 3 and made it easier to detect categories; 14 categories were thus found. In this analysis, based on the number and combination of extracted words with ties and the values of the coefficients, the determination was that interpreting the data in five categories, from 01 to 05, was desirable. In addition, because the Jaccard coefficient has the problem that the more frequent a word is, the less related it is to other extracted words, the author checked the most frequent values of extracted words and found that “write (書く)” was used 161 times, followed by “think (思う)” 101 times, and “myself (自分)” 95 times. In the co-occurrence network depicted in Figure 3, “write” and “think” were classified in the same category and were meaningful expressions; the present Jaccard coefficient was interpreted as being supportive.

A description of each category and an excerpt from the SREF are provided in the following. Category 01 is a collection of “effective use of conjunctions.”

In the beginning (初めは), the sentences were contradictory and didn’t convey what I wanted to say, but as I wrote more and more reports, I thought about conjunctions (接続詞) and the order in which they should be stated, and I think I gained the ability to notice (気付く) contradictions. (Student 69).

Category 02 is “appropriate content understanding and application.”

At first, I didn’t know how to write my thoughts, and my sentences became confused, mixing facts and my thoughts, or I couldn’t take in the content (内容) of the class and wrote something different from the fact. I think I was able to understand (理解する) and summarize the content (内容) little by little through the three upgrades. (Student 123)

Category 03 is “appropriate organization of figures, tables, and paragraphs.”

The first time, I wrote sentences mostly based on my own speculations, so it could not be called a report; the second time, my opinion was different in each paragraph (段落) of the sentences I wrote, and in the end, I could not understand what I wanted to claim. This time, I was able to write my argument first and make it clear. (Student 4)

Category 04 is “logical explanation of solutions to social problems.”

The first time I wrote it, I was focusing on my own ideas, so my logic was contradictory and leaps and bounds, but after writing it three times, I became aware of it. To explain (説明する) my evidence, I was able to learn about the problems (問題) that modern society (社会) is facing and the solutions to those problems (問題). (Student 105)

Category 05 is “constructing persuasive evidence and opinions.”

I have improved my writing (書く) skills and am able to write (書ける) logically. I have gained the ability to look for evidence (論拠) to make my opinions (意見) more persuasive. I realized that it is important to make a framework before writing (書く) sentences (文章), and by making a framework, I can write (書く) sentences (文章) that are easier to understand (分かりやすい) than before. (Student 109)

Teacher Aizawa also answered the question in the written interview, “How do you think the students have changed in terms of their abilities and attitudes through this practice?” He responded as follows.

Before I taught them, they just wrote their own arguments as they thought of them, and they were extremely poor in evidence. The writing style was not logically structured or expressed, and many of them were like “long tweets.” After instruction, the students showed an awareness of presenting objective, quantitative, and qualitative arguments (information-seeking skills) and developing their own arguments (problem-solving skills), and the structure of their writing was based on introduction, body, and conclusion (logical structure skills). Inappropriate colloquial vocabulary was reduced (language skills). This was noticeable after giving feedback to each student with detailed comments. (Parentheses are added by the author)

Therefore, from the first draft to the third draft, the following changes were made as a learning outcome of the ipsative assessment: from “subjective explanations” that were not logical to “logical structure skills” aware of “effective use of conjunctions” and “appropriate organization of figures, tables, and paragraphs”; from immature and inaccurate language to “language skills” of “objective expressions” and “appropriate understanding and application of content”; from abstract arguments to “problem-solving skills” to striving for “logical explanations of solutions to social problems”; and the “information-seeking skills” to search for evidence from various perspectives and conduct “detailed analysis” to “construct persuasive evidence and opinions” from analysis based on “thin evidence” by using only given materials. The continuous, cumulative formative feedback contributed to the overall improvement in writing skills by using inquiry learning. Next, an analysis is presented of the characteristics of formative feedback factors that encouraged students to improve their reports.

Opening Assessment Criteria

First, the teacher’s formative feedback (IFB1 and IFF2) on the second draft reports for 137 students was organized on the basis of FAT. Teacher Aizawa assessed 1,158 instances of formative feedback (including duplicates), indicating which criterion of the rubric they corresponded to, and communicated
them to the students. After the experiment, the author reclassified the 1,158 comments into FAT. The results are shown in Table 6. For example, 135 of the A1s were in category p1, with A1 accounting for 44.41% of all p1 and p1 accounting for 63.38% of all A1s. Additionally, 213 A1 and 304 p1 accounted for 18.39% and 26.25% of all formative feedback, respectively.

Based on Table 6, Teacher Aizawa gave a lot of praise (p1) for A1, A2, and C. He judged them to be “well done” in terms of appropriate and critical reading of the material and application of previously learned material.

This is an excellent perspective. It does not simplify the discussion but tries to capture the reality from multiple perspectives. (A2-p1)

In B1 and B2, many comments were written on approach critiques of structure and argument (c3), specific points (a1), and requests for explanation and questions (q), accounting for nearly 80% of the total comments in B1 and B2, respectively.

What does it mean to “deepen understanding” or “take a step forward?” It’s very abstract. (B1-q)

The descriptions are not connected, making it difficult to read. I think it would be better if you first state what you want to say straightforwardly. (B2-a2)

In addition to p1, the formative feedback that Teacher Aizawa considered C “integrated knowledge management” was concentrated in the factual critiques of the content (c2). There was a noticeable misunderstanding or lack of understanding in knowledge and content. Examples of c2 comments are as follows.

I think you have the image that capitalism is a society of disparity and socialism is a society of (economic) equality. Is this really so? There are many capitalist countries in Europe with small disparity, and there are (and have been) socialist countries with large disparity. (C-c2)

Most of the comments in D “writing style” were corrections of errors (c1), which is a logical and reasonable classification. The teacher used the criteria in D to point out c1 “corrections of errors.” In addition to mispronunciation, the teacher pointed out errors; for example,

This is the only part of the paper that is not written in the “desu-masu” end-of-sentence phrase style. Unify the “da-dearu” style. (D-c1)

Next, the lines of FAT were examined, and there was bias in the formative feedback provided by Teacher Aizawa. No unclassified (O) content was identified. In order of decreasing number, there were few remarks for future tasks (a3), recognition of progress (p2), and general remarks (a2), supporting the fact that p2 and a3 were not frequent in Hughes et al.’s (2015) study. In particular, students were asked to conduct the IFB themselves; thus, the teacher’s interest in p2 may have been further diminished.

No matter what kind of essay you encounter in the future, never use “I don’t know.” It is the same as submitting a test answer on a blank sheet of paper. (B1, B2, D-a3)

In c3, many comments were made on B1 and B2, mainly regarding inconsistencies and leaps in logic, and ambiguity and immaturity of claims and arguments. a1 and q had a high ratio of B1 and B2. For example,

This is also very twisted and difficult to understand. You say, “We should do something to close the gap,7” but here you change 180 degrees to “We don’t necessarily need to close the gap.” Which is your point? It’s a contradiction. (B1, B2-c3)

That is exactly what this university is asking. If you think that we should try to prevent the widening of the gap, as in the case of universities, then I think you need to discuss your ideas on how to prevent it. (B2-a1)

In this paragraph, you seem to define trade and exchange as “something that takes place from the rich to the poor.” Is this perception really valid? Is it really complementary? (B1, B2-q)

Notably, Teacher Aizawa answered that in some cases, multiple rubric criteria are applicable. In other words, by crossing and blurring the existing criteria, formative feedback was implemented in reference to a newly transformed criterion, that is, the ipsative criterion, which was not explained by the rubric criteria.

Thus, the next step was to analyze the content of the SIFB to identify the criterion. The total comments for SIFB1 and SIFB2 were 113 and 425, respectively. The increase in the number of comments was due to the higher number of TIFB1 and TIFF2 comments on the second draft report. The actual numbers (on the bar graph) and percentages were calculated in correspondence to each criterion in the rubric; the results are shown in Figure 4.

In total, the SIFB interests were categorized into 22 criteria. In those, 17 criteria, A1 “accurate reading of materials and questions”; A2 “critical reading of materials and questions”; B1-1 “logical structure”; B2-2 “analysis of events”; C1 “application of previous knowledge”; D1 “sentence form”; D2 “appropriate words”; D3 “easy-to-understand style”; D4 “paragraph structure”; D5 “agreement of main statements”; and E1 “concretization of the whole description” and E2 “critical eye,” which did not fit into any of the criteria; were common to SIFB1 and SIFB2. B1-6 “reinforcing arguments,” B1-7 “revising arguments,” B2-3 “multiple evidence,” B2-4 “rebuttal,” and C2 “selecting appropriate knowledge” were found only in SIFB2 and were new criteria for students by TIFB1 and TIFF2.
Table 6. Typology of TIFB1 and TIFF2 (Cross Table of the Rubric and FAT).

| Rubric | A1 | A2 | B1 | B2 | C   | D   | Frequency | Total | Rank |
|--------|----|----|----|----|-----|-----|-----------|-------|------|
| FAT    |    |    |    |    |     |     |           |       |      |
| p1     | 135| 44.41% | 42 | 13.82% | 34 | 11.18% | 27 | 8.88% | 50 | 16.45% | 16 | 5.26% | 26.25% | 304 | 1 |
| p2     | 63.38% | 67.74% | 12.19% | 11.64% | 45.45% | 6.11% | 60.00% | 0.43% | 5 | 8 |
| c1     | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 1.15% |
| c2     | 39 | 38.24% | 1 | 0.98% | 13 | 12.75% | 17 | 16.67% | 31 | 30.39% | 1 | 0.98% | 8.81% | 102 | 6 |
| c3     | 18.31% | 15.61% | 1 | 0.58% | 58 | 33.53% | 70 | 40.46% | 16 | 9.25% | 1 | 0.58% | 14.94% | 173 | 3 |
| a1     | 6 | 3.66% | 12 | 7.32% | 77 | 46.95% | 56 | 34.15% | 5 | 3.05% | 8 | 4.88% | 14.16% | 164 | 5 |
| a2     | 2.82% | 19.35% | 27.60% | 24.14% | 4.55% | 3.05% | 18 | 50.00% | 3.11% | 36 | 7 |
| a3     | 0% | 0% | 0% | 0% | 0.36% | 0.43% | 0% | 0.38% | 3 | 9 |
| q      | 6 | 3.61% | 6 | 3.61% | 86 | 51.81% | 51 | 30.72% | 8 | 4.82% | 9 | 5.42% | 14.34% | 166 | 4 |
| o      | 2.82% | 9.68% | 30.82% | 21.98% | 7.27% | 3.44% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 10 |

Frequency: 18.39% 5.35% 24.09% 20.03% 9.50% 22.63% 100%
Total: 213 62 279 232 110 262 1158
The low rate of change in A1 and A2 among SIFB1 and SIFB2 was presumably due to the high rate of p1 in TIFB1 and TIFB2, which reduced student interest. Although the number of comments on D, which also declined, was high, there was more interest in the other criteria. B1 and B2 increased from 23.01% to 39.53% and from 30.97% to 56.24%, respectively, and were the criteria with increasing rates of change. Therefore, the students interpreted and transformed the criteria in the rubric to suit their learning but seemed to be more interested in improving the format of the report and other presentation aspects. Teacher Aizawa actively used the complex structure of the criteria in the formative feedback of a1, c2, c3, and q, especially in B1 and B2, to create five ipsative criteria for the claims and rationale of the drafts and their validity. The hallmark of ipsative assessment is not to match learning with a given goal or criteria but to align assessment and learning by allowing students to create educational values that best reflect their learning progress and difficulties.

Limitation of Students’ Self-Assessment

Regarding the reports and the SIFB, the second draft showed that the students still had problems, especially in logical organization and understanding and using knowledge. Teacher Aizawa viewed this result as a problem with the students’ assessment. He was keenly aware of the importance of peer assessment in his written interview.

The biggest takeaway (and failure) was that I found that the students’ self-assessment was not working very well; at the end of the third report, I included a section for them to write feedback to themselves. At the end of the third draft, there was a space for students to write feedback to themselves. Although I gave them examples of how to write it, it did not go over very well, and most of them wrote very shallow and abstract statements such as “I got better” or “I did my best.” The reason for this is that I have not educated my students on what it means to give deep feedback to themselves and others. I felt that it was necessary to incorporate ipsative assessment into daily learning many times, and to improve the quality of “deep feedback” by sharing it with students and teachers.

However, Teacher Urakawa, who attended the conference, also sensitively reported that his students’ second drafts were poor because of attempting peer assessment (his class was not analyzed in this study). Sadler (1989) stated that students have to understand the criteria that exhibit characteristics of the almost same quality as the teacher for them to adjust to learning. Self-assessment and peer assessment should be encouraged, but their outcomes depend on students’ level of assessment skills.

The limitations in the students’ improvement can also be analyzed in terms of how much they revised their reports. Therefore, the writing content changes from the first draft to the second draft (1→2 group) and from the second draft to the third draft (2→3 group) were examined using the sentence concordance percentages in 126 cleaned reports. The average of each concordance rate for the 1→2 group is 22.2% (median = 19.6%, maximum = 84.0%, minimum = 0%), and that for the 2→3 group is 6.59% (median = 2.6%, maximum = 51.7%, minimum = 0%). To observe if the percentage change between the two groups was statistically significant, the author performed the following analysis methods by using Easy R version 1.54. The proportions were inverse sine transformed to stabilize the variance, and then a corresponding t-test was performed between the 1→2 and 2→3 groups. The angular transformed values were determined for the 1→2 group (mean = 0.450, standard deviation = 0.242) and the 2→3 group (mean = 0.195,
Challenges and Solution Proposals

Finally, the author confirms the kinds of challenges that emerged when existing theoretical and methodological frameworks of ipsative assessment were applied to secondary education in Japan and then provides indications as to how they can be addressed.

The first challenge was the burden on teachers of assessment activities. In a written interview, Teacher Aizawa described that “the size of one class (40 students) was too large, and writing feedback and comments was hard work.” SIFB1 and SIFF2 were administered during the summer vacation, when there were no classes, but devoting time to assessment during daily classes is difficult. Although some prior research (e.g., Hughes, 2014; Malecka et al., 2021) proposed that technology could efficiently organize assessment data, establishing online learning environments in Japanese classrooms was not possible because of the lack of policy budgets, and teachers’ positive reliance on the traditional educational system. The more pressing issue is that school organizations need to take the responsibility to balance the ratio of assessment activities and overall workload. Therefore, in the 2021 school year, the School increased the number of “Contemporary Society” teachers in charge to five, with one teacher per class, and eliminated the regular examinations for “Contemporary Society,” developing a curriculum that allows students to focus on inquiry learning and ipsative assessment year-round.

The second challenge was the development of the students’ assessment skills. As described in Limitation of Students’ Self-Assessment, the students’ self-assessment only worked well when there was supportive management. Seery et al. (2019) showed in their experimental paradigm based on adaptive comparative judgments that only by repeatedly examining the evidence presented by their peers were students able to enrich their evaluative judgments and construct standards and criteria. Malecka et al. (2021) found that the ipsative use of e-portfolios encouraged students to access their peers’ work easily and that timely feedback and feedforward were useful in maintaining positive motivation. A collaborative learning environment is required for students to connect assessment to their learning. Since the 2020 school year, the School’s teachers and students had been collaboratively constructing educational values, such as by offering a class during which high school students read reports written by university students from the School, philosophically perform a dialog on what criteria define a good report, and learn to understand the concept of quality.

The third challenge was the development of teachers’ assessment skills. The results of 149 SNEGs completed by 91 students were categorized (Table 7). Notably, despite the formative feedback by Teacher Aizawa, the students had many diverse problems. One-time formative feedback does not solve the problems, and that formative feedback creates new learning difficulties. Teachers should continue to base IFB and IFF on the dialog with students rather than doing so on learning products. Research and support have to focus on teachers’ perceptions of what proper IFF and IFB are to ensure students’ personal best. As Malecka and Boud (2021) mentioned, autoethnographic research could reveal how teachers perceive assessment and operationalize an ipsative approach to curriculum. Just as importantly, intervention from researchers to support proficiency in formative assessment practices is essential (Black & Wiliam, 1998). However, some things need to be considered for interventions. While various programs have been conducted that develop formative assessment knowledge and skills (Andersson & Palm, 2017; Fischman & Wasserman, 2017; Hopfenbeck et al., 2012), these have been found to have contextual compatibility and sustainability problems (Black & Wiliam, 2018; DeLuca et al., 2012; Yan et al., 2021). Researchers and school supporters must support teachers struggling against authoritative political and social structure discourses in a bottom-up manner to enable them to imbibe learning that allows them to develop formative assessment concepts on the basis of their culturally optimal ways of thinking and doing.

The fourth challenge was students’ intrinsic or extrinsic motivation to learn. In the oral interviews, Teacher Aizawa and Urakawa stated that students felt bored with repeating the same task. To motivate students to learn, the quality and subject matter of the assessment are important. In this study, the teachers asked the students to work on a predetermined topic that some students might not have been interested in. This case study, which used existing theoretical and methodological frameworks, has demonstrated that previous ipsative assessment practices have not been overly concerned with tasks or task-related curricula. Therefore, the School is attempting to inform students of the value of learning by arranging the curriculum so that

standard deviation = 0.181). The result is $p = 2.63e-21 < 0.01$; thus, the percentage of students who copied-and-pasted was significantly higher in the 2→3 group compared with the 1→2 group. Students changed the content of their report writing more significantly in the second improvement than in the first improvement.

Thus, the content-level assessment by the teacher (improvement from the second to the third draft) was more effective than the format-level self-assessment by the students (improvement from the first to the second draft). Notably, the rubrics used in this study were not specific but rather higher-order criteria, increasing the students’ difficulty in understanding the criteria, which may have been a factor in narrowing the improvement activities. For ipsative assessment in large classrooms, it is necessary to consider the curriculum that makes the assessment itself a learning content, and the system (e.g., use of information and communication technology equipment) for conducting the assessment is required.
students can learn inquiry-based academic writing in the first and second semesters and apply it in the “open inquiry” in the third semester.

The fifth challenge was an explanation of the relationship between summative ipsative assessment and criterion-referenced assessment. Teacher Aizawa scored the third draft by using a rubric. Although the method of this study is compatible with criterion- and self-referenced assessments, the absence of standardization of the assessment results may be problematic in the accountability of grades. In this study, the theoretical ipsative assessment framework by Hughes (2014) was adopted; however, it did not seem complete. The debate should continue on whether criterion-referenced and ipsative assessment should each be performed in isolation or whether they should be harmonized. Therefore, this study adopted an experimental design that was more focused on this harmonization. Given that the epistemologies for criterion-referenced assessments and ipsative assessments are fundamentally different (Table 1), simplistic claims (e.g., Harlen, 2012) that both follow the same process should be treated with caution because how the two positions are merged and differentiated can vary depending on the meaning of the summative assessment and learning. Therefore, it could be concluded that the theoretical framework for ipsative assessment is still developing. A rigorous experimental design, including a classroom-level summative assessment, needs to be prepared after clarifying the orientation of the ipsative assessment.

### Conclusion

In this study, the author discussed the significance and challenges of formative ipsative assessment from the view of inquiry learning in a Japanese high school. Quantitative and qualitative data from multiple perspectives were collected to attempt an integrated analysis and discussion. The results of three rounds of academic writing for high school freshmen showed that the practice designed in this study gradually improved students’ overall writing skills; thus, it is expected to result in qualitative improvement and transformation in performance-centered learning. In Japanese high schools, where a single teacher is responsible for many students and is constrained by external standards, one strategy is to align learning and assessment by developing abstract higher-order criteria and then developing them into concrete lower-order criteria that are appropriate for each student. Of particular importance is that (1) students should work on the same task that maintains their motivation to learn, (2) the solution to the task should be essential and not fixed, (3) students should reflect specifically on what has improved or where progress has not been made compared with past learning, (4) students should apply the improvements to their next learning, (5) students should compare their first and last products and reflect on “growth as transformation,” (6) teachers should provide formative feedback cumulatively and continuously, (7) teachers should timely deal with students’ learning needs, and (8) teachers should embed formative ipsative assessment practices in the curriculum to allow them to run parallel to daily teaching.

### Table 7. Contents of SNEG.

| Category                        | Comment examples                                                                 | No. |
|---------------------------------|----------------------------------------------------------------------------------|-----|
| Forming opinions                | I was unsure if my idea was realistic and if it was worth arguing for.           | 17  |
| Logic configuration             | I tried not to make any logical leaps, but I couldn’t understand it.             | 16  |
| Knowledge                       | It was difficult to read the material due to my lack of social knowledge.        | 15  |
| Organizing thoughts             | I ended up saying the same thing over and over again, and I didn’t know what I wanted to say. | 14  |
| Expressive skills               | I could not express what I wanted to say well.                                  | 13  |
| Making connections between sentences |                                                                                | 12  |
| Finding appropriate evidence    | I could not find the references I was looking for.                             | 11  |
| Report writing techniques       | When I write a report and don’t have access to a computer and data, how do I write the evidence? | 11  |
| Forming evidence                | Not being able to determine from where to what part of the report an argument is needed. | 7   |
| How to write a conclusion       | How concise should the conclusion be written?                                   | 6   |
| How much to write               | If I try to divide a paragraph into “Introduction, Main Body, and Conclusion,” I end up with a long paragraph. In general, how many lines should a paragraph have? | 5   |
| Combining facts and opinions    | I couldn’t connect the material and my opinion well even though I was stating my opinion based on the material. | 4   |
| Rebuttals                       | I didn’t know how to write my opinion on the counterargument.                   | 4   |
| Criticism                       | I would like to know how well everyone else is able to write.                   | 3   |
| Learning improvement            | Even though I tried to improve my writing, I didn’t know how and where to improve. | 3   |
| Meaning of criteria             | I still don’t understand what logic is.                                         | 2   |
| Others                          | Lack of time. How to title, etc.                                                | 4   |
The limitations of this study are twofold. First, the experimental design was incomplete. As all the students received the same TIFF1, tracking the transformations of the quantitative and qualitative characteristics of the categorical comments in the FAT between TIFF1, and TIFF2 and TIFF1 was not possible. Thus, further research should be conducted with a more rigorous experimental design. Further research should establish a better practical methodology by implementing it in a broader context of educational goals and objectives and in various learning content areas, with sufficient discussion of the links with summative assessment.

Second, the context-dependent nature of ipsative assessment theory was not considered. Although ipsative assessment and FAT are based on feedback research in the field of higher education, they have also been applied in primary and secondary education and are expected to be effective (Hughes, 2017). However, for secondary school students, who may be concerned about scores, the “score justification” category eliminated during the FAT creation phase may be important (Hughes et al., 2015). In addition, the epistemology of ipsative assessment is rooted in artistic ideas that are close to the “progress of excellence” explored by Budo (Japanese martial arts) and Geido (Japanese accomplishments) in terms of “qualitative improvement of activities across the lifespan.” Therefore, FAT alone may overlook the community-specific aspects of formative feedback and the possibility that each learning domain has unique characteristics. More exploratory data collection and analysis on the concept and types of assessment should be continued to accumulate additional ipsative practices. Specific case studies from culturally developed places are also needed to expand the possibilities and identify the limitations of the theory, such as in schools with strong Japanese cultural characteristics or in classes that allow high-quality student reflections. Ipsative assessment is a concept that should be theorized within an interdisciplinary field, not discussed only in the schooling dimension.

Declaration of Conflicting Interests
The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by JSPS KAKENHI Grant Number 20J20092.

ORCID iD
Kohei Nishizuka https://orcid.org/0000-0002-7488-3745

Supplemental Material
Supplemental material for this article is available online.

Notes
1. In this article, assessment means the process of observing, collecting, and organizing information on the status and progress of learning.

2. This research design was approved by the Ethics Committee at Graduate School of Education, Tohoku University, Japan (Approval ID: 19-1-009).

3. The trading game is a simulation game in which a country’s resources and technology are replaced by paper and tools (e.g., pencils, scissors), and each assigned nation (team) competes to see how much wealth it can acquire by making products (e.g., rectangular paper) by using different resources and technology and negotiations among nations.

4. The inequality “r (return on capital) > g (economic growth rate)” is a disparity factor proposed by Thomas Piketty.

References
Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Education & Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017
Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. *Education Inquiry, 8*(2), 104–122. https://doi.org/10.1080/20004508.2016.1275185
Azizifard, N. (2014). Social network clustering. *International Journal of Information Technology and Computer Science, 6*(1), 76–81. https://doi.org/10.5815/ijitcs.2014.01.09
Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children, 46*(2), 26–29.
Barron, B., & Darling-Hammond, L. (2010). Prospects and challenges for inquiry-based approaches to learning. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 199–225). OECD Publishing.
Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education Principles Policy and Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102
Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education Principles Policy and Practice, 26*(6), 551–575. https://doi.org/10.1080/09695959x.2018.1441807
Chen, C. H., & Yang, Y. C. (2019). Revisiting the effects of project-based learning on students’ academic achievement: A meta-analysis investigating moderators. *Review of Educational Research, 26*(1), 71–81. https://doi.org/10.5815/ijitcs.2014.01.09
Christodoulou, D. (2017). Making good progress? *The future of assessment for learning*. Oxford University Press.
Conway, C. J. (2014). Effects of guided inquiry versus lecture instruction on final grade distribution in a one-semester organic and biochemistry course. *Journal of Chemical Education, 91*(4), 480–483. https://doi.org/10.1021/ed300137z
DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters, 4*, 5–29. https://doi.org/10.18296/am.0104
Elwood, J. (2006). Formative assessment: Possibilities, boundaries and limitations. *Assessment in Education Principles Policy and Practice, 13*(2), 215–232. https://doi.org/10.1080/0969595940600708653
Engeström, Y. (2015). *Learning by expanding: An activity-theoretical approach to developmental research* (2nd ed.). Cambridge University Press.
Fischman, D., & Wasserman, K. (2017). Developing Assessment through Lesson Study. *Mathematics Teaching in the Middle
Hughes, G., Smith, H., & Creese, B. (2015). Not seeing the wood
Ipsative assessment and personal learning gain:
Hopfenbeck, T. N., Throndsen, I., Lie, S., & Dale, E. L. (2012).
Higuchi, K. (2019). Using correspondence analysis in a quantita-
Hattie, J., & Timperley, H. (2007). The power of feedback.
Visible learning: A synthesis of over 800 meta-
Harlen, W. (2012). On the relationship between assessment for
Harlen, W. (2003).
Grob, R., Holmeier, M., & Labudde, P. (2017). Formative assess-
Fleer, M. (2015). Developing an
Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012).
Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. Assessment in Education Principles Policy and Practice, 14(3), 387–392. https://doi.org/10.1080/09695940701592097
Sadler, D. R. (1987). Specifying and promulgating achievement standards. Oxford Review of Education, 13(2), 191–209. https://doi.org/10.1080/0305498870130207
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18(2), 119–144. https://doi.org/10.1007/bf00117714
Penn, P. R., & Wells, I. G. (2018). Making assessment promote effective learning practices: An example of ipsative assessment from the School of Psychology at UEL. Psychology Teaching Review, 24(2), 70–74. https://doi.org/10.1177/1475725718814315
Rattray, J. (2018). Affect and ipsative approaches as a counter to pedagogic frailty: The guardian of traditional models of student success. Higher Education Research & Development, 37(7), 1489–1500. https://doi.org/10.1080/07294360.2018.1494141
Malecka, B., & Boud, D. (2021). Fostering student motivation and engagement with feedback through ipsative processes. Teaching in Higher Education, 1–16. https://doi.org/10.1080/13562517.2021.1928061
Malecka, B., Ajjawi, R., Boud, D., & Tai, J. (2021). An empirical study of student action from ipsative design of feedback processes. Assessment & Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602933.2021.1968338
Margunayasa, I. G., Dantes, N., Marhaeni, A. A., & Suastra, I. W. (2019). The effect of guided inquiry learning and cognitive style on science learning achievement. International Journal of Instruction, 12(1), 737–750. https://doi.org/10.29333/iji.2019.12147a
Nishizuka, K. (2020). A critical review of formative assessment research and practice in Japan. International Journal of Curriculum Development and Practice, 22(1), 15–47. https://doi.org/10.18993/jcrdaen.22.1_15
Nishizuka, K., & Arimoto, M. (2019). Reconsidering the contemporary significance of assessment criteria on the theory of formative assessment: Toward the implementation of “deep education for sustainable development.” Studies in Holistic Education Care, 22, 45–60. (In Japanese)
Pedaste, M., Mäeots, M., Sitman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. Review of Educational Research, 44, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003
Penn, P. R., & Wells, I. G. (2018). Making assessment promote effective learning practices: An example of ipsative assessment from the School of Psychology at UEL. Psychology Teaching Review, 24(2), 70–74. https://doi.org/10.1177/1475725718814315
Rattray, J. (2018). Affect and ipsative approaches as a counter to pedagogic frailty: The guardian of traditional models of student success. Higher Education Research & Development, 37(7), 1489–1500. https://doi.org/10.1080/07294360.2018.1494141
Sadler, D. R. (1987). Specifying and promulgating achievement standards. Oxford Review of Education, 13(2), 191–209. https://doi.org/10.1080/0305498870130207
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18(2), 119–144. https://doi.org/10.1007/bf00117714
Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. Assessment in Education Principles Policy and Practice, 14(3), 387–392. https://doi.org/10.1080/09695940701592097

Hughes, G. (2017). Ipsative assessment: Motivation through marking progress: Palgrave Macmillan.
Hughes, G. (2017). Ipsative assessment and personal learning gain: Exploring international case studies. Palgrave Macmillan.
Hughes, G., Smith, H., & Creese, B. (2015). Not seeing the wood
Ipsative assessment and personal learning gain:
Hopfenbeck, T. N., Thronsden, I., Lie, S., & Dale, E. L. (2012).
Assessment with distinctly defined criteria: A research study of a National Project. Policy Futures in Education, 10(4), 421–433. https://doi.org/10.2304/pfie.2012.10.4.21
Hughes, G. (2014). Ipsative assessment: Motivation through marking progress: Palgrave Macmillan.
Higuchi, K. (2019). Using correspondence analysis in a quantitative text analysis procedure. Computers & Education, 47, 18–24. (In Japanese) https://doi.org/10.14949/konpyutariyoukyouiku.47.18
Higuchi, K. (2020). Quantitative text analysis for social researchers: A contribution to content analysis (2nd ed.). Nakanishiya Shuppan. (In Japanese)
Hopfenbeck, T. N., Thronsden, I., Lie, S., & Dale, E. L. (2012). Assessment with distinctly defined criteria: A research study of a National Project. Policy Futures in Education, 10(4), 421–433. https://doi.org/10.2304/pfie.2012.10.4.21
Hughes, G. (2014). Ipsative assessment: Motivation through marking progress: Palgrave Macmillan.
Hughes, G. (2017). Ipsative assessment and personal learning gain: Exploring international case studies. Palgrave Macmillan.
Hughes, G., Smith, H., & Creese, B. (2015). Not seeing the wood
Ipsative assessment and personal learning gain:
Hopfenbeck, T. N., Throndsen, I., Lie, S., & Dale, E. L. (2012).
Assessment with distinctly defined criteria: A research study of a National Project. Policy Futures in Education, 10(4), 421–433. https://doi.org/10.2304/pfie.2012.10.4.21
Hughes, G. (2014). Ipsative assessment: Motivation through marking progress: Palgrave Macmillan.
Higuchi, K. (2019). Using correspondence analysis in a quantitative text analysis procedure. Computers & Education, 47, 18–24. (In Japanese) https://doi.org/10.14949/konpyutariyoukyouiku.47.18
Higuchi, K. (2020). Quantitative text analysis for social researchers: A contribution to content analysis (2nd ed.). Nakanishiya Shuppan. (In Japanese)
Hopfenbeck, T. N., Thronsden, I., Lie, S., & Dale, E. L. (2012). Assessment with distinctly defined criteria: A research study of a National Project. Policy Futures in Education, 10(4), 421–433. https://doi.org/10.2304/pfie.2012.10.4.21
Hughes, G. (2014). Ipsative assessment: Motivation through marking progress: Palgrave Macmillan.
Seery, N., Buckley, J., Delahunty, T., & Canty, D. (2019). Integrating learners into the assessment process using adaptive comparative judgement with an ipsative approach to identifying competence based gains relative to student ability levels. *International Journal of Technology and Design Education, 29*, 701–715. https://doi.org/10.1007/s10798-018-9468-x

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*(3), 323–342. https://doi.org/10.1080/03054985.2012.689693

Univio, D. J., & Pérez, A. D. (2019). Ipsative assessment of essay writing to Foster Reflection and self-awareness of Progress. In E. White & T. Delaney (Eds.), *Handbook of Research on Assessment Literacy and Teacher-Made Testing in the Language Classroom* (pp. 157–180). IGI Global.

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers’ intentions and implementations regarding formative assessment. *Assessment in Education Principles Policy and Practice, 28*(3), 228–260. https://doi.org/10.1080/0969594x.2021.1884042