

An Evaluation Framework Connecting Product State, Provision, and Eating Experience

Conceptual Structure and Operational Principles of ICES, SEAS, and TDC

Toshiharu Kaneko

Cheese Art Fromager Association of Japan (CAFAJ), Japan

28 September 2026

Abstract

Cheese evaluation requires product-state conformity to specifications, the suitability of provision for its intended recipients and purpose, and the sensations and preferences experienced by the eater to be considered with their respective evidence. This paper examines the International Cheese Evaluation Standard (ICES), Sensory Evaluation Applied Standard (SEAS), and Three-Dimensional Composition Theory (TDC), systems developed from knowledge accumulated through the author's twelve years of cheese education practice and relevant prior research. It presents product state, provision conformity, and composition–experience relationships as a conceptual model. Comparison with research on cheese sensory vocabulary, presentation conditions, dairy-product specification compliance, meal services, and quality perception clarifies the relationships between two references—product specifications and provision goals—and the individual's experiential outcomes. The unit of connection is an observation record identifying the sample, conditions, time, and item, retaining the principles of evidence referencing and independent judgment. Four testable propositions address judgment traceability, shared reference specifications, manipulation of composition, and independently measured goal attainment. The contribution is the formalization of practice-derived knowledge as a judgment structure available for external reference and examination. The analysis addresses concepts and operational structure. Research on specific effects and applicability will continue through participant-response analyses and comparative studies.

Keywords: cheese; sensory evaluation; provision design; multisensory perception; evaluation framework; vocational education

1 Background and purpose

Judgments that the same cheese is in good condition, appropriately provided, or liked answer different questions. Product-state diagnosis requires observations relative to a reference specification established in advance. Provision design requires judgments that consider recipients, purpose, setting, and operations. Understanding eating experience requires records of what the eater perceives, does, and evaluates. Recording these judgments and their evidence separately clarifies which conditions to examine for improvement.

This paper examines three systems grounded in the author's cheese education and provision practice. The International Cheese Evaluation Standard (ICES) addresses product state; the Sensory Evaluation Applied Standard (SEAS) addresses provision conformity for intended recipients and purposes; and Three-Dimensional Composition Theory (TDC) addresses relationships between composition, perceptual and eating processes, and eating experience [1]. The framework draws on provision and

teaching knowledge accumulated through the author's twelve years of cheese education practice and relevant prior research. This period denotes the accumulation of practical knowledge. The paper presents the systems' definitions, judgment procedures, and recording conditions as a conceptual model.

The research question is how product state, provision conformity, and eating experience can be connected for education and improvement while retaining their distinct meanings. Specifically, the paper defines the objects of observation and judgment references of the three systems, identifies information that can be shared and numerical values that cannot be transferred between them, and derives testable propositions about the usefulness of their connection. This provides a route from experience-based judgment to records that others can retrace.

2 Relationship to prior research

2.1 Cheese description and shared understanding among assessors

Drake and colleagues [3] developed a shared descriptive vocabulary for Cheddar cheese flavor and examined whether a vocabulary supported by references could distinguish samples. Their subsequent study [4] compared evaluations by trained panels at multiple sites and addressed the practical value of shared definitions and references. These studies provide a research foundation for the use of terminology, references, training, and evaluation procedures to articulate and share sensory experience.

In addition to describing attribute intensity, ICES compares product state with a reference specification defined in advance. Scoring dairy-product sensory properties against product specifications is also addressed in ISO 22935-3:2023 [13]. Building on established vocabulary development and specification-conformity evaluation, this paper specifies the conditions under which ICES domain findings can inform purpose-specific provision judgments. The validity of the particular ICES item structure is evaluated with reference to those items and their conditions of application.

2.2 Presentation conditions and eating experience

Michel and colleagues [5] presented the same ingredients in different visual compositions and examined effects on evaluations before and after eating. Harrar and Spence [6] manipulated cutlery weight, size, color, and shape to investigate changes in food perception and evaluation. These experiments show that presentation conditions beyond the food itself can contribute to eating experience. Interpretation specifies each study's scope in terms of participants, foods, and presentation conditions. For Michel and colleagues' study, possible differences in ingredient contact and mixing accompanying arrangement are also considered alongside visual conditions.

TDC distinguishes manipulable composition, perceptual and eating processes, and evaluative outcomes so that these relationships can be recorded and examined in cheese provision practice. Building on existing research on composition and eating experience, this paper sets out a method of connecting those relationships, through observation

records, to product-state diagnosis and provision judgments relative to purpose.

2.3 Sensory evaluation standards and the scope of this paper

ISO 6658:2017, published by the International Organization for Standardization (ISO), provides general guidance on sensory evaluation methods, while ISO 8586:2023 addresses assessor selection and training [7,8]. ISO 11132:2021 addresses quantitative descriptive panel performance, including discrimination, panel agreement, and repeatability of intensity ratings [9]. This paper consults the official public summaries to compare the scope and methodological positioning of these standards.

ISO 11132 concerns quantitative descriptive panels. Methods for evaluating the four-level conformity judgments of ICES and SEAS are selected to match the scales and research purposes. Reproducibility across sites or different conditions is addressed through designs appropriate to each question. International standards are consulted as a foundation for shared procedures, while the effectiveness of the item structures and indices considered here is evaluated through research directed at those systems.

2.4 Integrated models of meal services and quality perception

The Five Aspects Meal Model (FAMM) of Gustafsson and colleagues considers meal services in terms of room, meeting, product, management control system, and atmosphere [14]. This paper draws on that perspective encompassing food, people, and environment. FAMM organizes aspects of meal services, while SEAS organizes axes for judging conformity of provision to purpose. Comparison therefore considers their respective objects of classification and roles.

In a framework of food-quality perception, Grunert distinguishes quality expectations before purchase from quality experienced after purchase when discussing consumer food choice [15]. This paper likewise distinguishes expectation from actual experience and further organizes product-specification conformity, provision-goal conformity, and individual experiential outcomes into mutually referable records retaining their respective judgment evidence. Section 10 summarizes the positioning within the comparisons undertaken.

3 Materials and methods of conceptual analysis

3.1 Source materials and selection

The analysis uses the common-foundation textbook for fromager certification [1], CAFAJ's theoretical source for Three-Dimensional Composition Theory [10], and the specifications for the Sensory Evaluation Applied Standard and International Cheese Evaluation Standard [11,12]. Definitions, item structures, judgment statements, formulas, and recording procedures were analyzed to identify judgment references and conditions for connecting the systems. Definitions, judgment conditions, and formulas needed to understand the conceptual model and index calculations are presented in Sections 4–6.

External sources were selected for the purpose of the paper from conceptual-paper methodology, cheese descriptive vocabulary, presentation-related changes in eating experience, sensory evaluation standards, and integrated models of meal services and consumer quality perception. ISO 22935-3:2023 was included for dairy-product specification compliance, and FAIM and Grunert's quality-perception framework as integrated models. Primary information from publishers, authors' institutions, and ISO was checked; the distinction between full texts, abstracts, and public summaries is stated with the references. Sources were selected purposively, and scholarly positioning is discussed within the scope of the literature consulted.

3.2 Analytical procedure

Following Jaakkola's emphasis on explicit research design in conceptual articles [2], analysis proceeded as follows. For each system, the object, input, reference, judgment, and output were extracted. Occurrences of the same terms or observations were then compared, separating product conformity from conformity to people and purposes. Shared observation records were identified and connection rules were constructed without numerical conversion between systems. Finally, structural coherence was distinguished from practical usefulness, and testable propositions were developed to examine the latter. Section 4 defines the three systems, Section 5 explains their connection through observation records, and

Section 6 sets out judgment and index rules [1,10–12]. Section 7 provides a hypothetical illustration, Section 8 presents four propositions for comparative study, and Section 9 outlines research design. The analysis focuses on concepts and operational structure, connecting participant-response analysis and effect estimation to separate practice-based studies.

3.3 Practical foundation and scope of analysis

As the practical foundation, the author has used a sequence in cheese-styling courses consisting of tasting the cheeses to be used, a lecture on styling effects, practical composition, a tasting session after composition, and an anonymous questionnaire. Examination of the relationship between provision conditions and experience, and organization of that knowledge into textbooks and operational documents, underpin the problem addressed here. This accumulated practice forms the background to the conceptual analysis. Comparative experiments under uniform conditions are conducted according to individual research designs.

The direct objects of analysis are the documents in Section 3.1. Knowledge derived from practice, findings of prior research, the conceptual organization presented here, and propositions for comparative study are distinguished. Analysis of course questionnaires is addressed in a practice report specifying question and response formats, periods covered, units of aggregation, and retained records.

4 Definitions and objects of judgment

The three systems can address the same cheese-provision setting, but their units of analysis and judgment references differ. TDC is an applied theory for explaining and examining composition–experience relationships; SEAS and ICES are criteria systems for judging conformity for their respective objects. The state conformity index (SCI) summarizes product-state judgments, and the provision conformity index (OCI) summarizes provision judgments. Their calculation conditions and interpretation are set out in Section 6. Practical use, conceptual structure, specific effects, and assessment performance are explained with their respective evidence and scope.

Table 1 Units of analysis and outputs

System	Principal unit of analysis	Reference for judgment or examination	Output
ICES	Sample and evaluation time	Prior product specification or reference profile	24 item judgments in six domains; state findings and SCI
SEAS	Provision plan or setting with specified recipients and purpose	Purpose and optimal range defined in advance	20 item judgments in five axes; provision findings and OCI
TDC	Composition and observation of a piece, plate, or provision space	Question, hypothesis, and comparison conditions	Records of composition, processes, and outcomes; examination of relationships

4.1 Product-state judgment with ICES

ICES stands for International Cheese Evaluation

Standard (国際チーズ鑑評基準). Its six domains are external state (EX), internal state (IN), texture and mouthfeel (TX), aroma (AR), flavor (FL), and maturation and overall coherence (MA), with four items per domain and 24 items in total. Findings at the time of evaluation are compared with a reference specification fixed before evaluation [1,12].

Reference specifications are considered in the order indicated in the textbook, including public product specifications, recognized category or contractual specifications, producers' specifications, maturation targets and dispatch criteria, and profiles approved in advance by an expert panel. If assessors revise the reference to fit their impressions after tasting, the meaning of a given score changes. The version and scope of the reference document are therefore recorded. ICES is a criteria system for judging product state and is distinct from an ISO standard.

ICES provides a descriptive mode that records observation terms and findings according to the status of the reference specification [12]. Domain indices and SCI are displayed when the reference specification is established and the required number of valid judgments is available. Item-level judgment conditions and the reference conditions for the evaluation as a whole are recorded separately.

4.2 Provision-conformity judgment with SEAS

SEAS stands for Sensory Evaluation Applied Standard (官能評価適用基準). It has five axes: recipient (O), state (ST), compatibility (C), structure (S), and method (M), with four items per axis and 20 items in total. The intended recipients and purpose are defined first;

provision design is then judged against an appropriate range for that purpose [1,11]. The optimal range encompasses conformity to purpose, including excess and insufficiency, and is defined for the recipients and setting.

For example, cross-section discrimination and ease of comparison are important when provision is intended for comparative learning, whereas bite size, combinations, and elapsed time play different roles when provision is part of a meal. SEAS use requires identification of who defined the optimal range and on what grounds. Professional conformity judgments and the eater's own satisfaction or liking are recorded separately.

4.3 Composition and experience in TDC

TDC stands for Three-Dimensional Composition

Theory (三次元構成理論). TDC organizes relationships in three layers: spatial structure; perceptual and eating processes; and sensory and experiential evaluative outcomes. Observation items are selected from seven groups totaling 53 items: spatial structure 14, vision 13, aroma 3, basic tastes 5, texture and oral sensations 10, sensory integration and temporal change 3, and purpose and experiential evaluation 5 [1,10].

The three layers identify analytical roles, while the seven groups classify objects of observation. The number 53 denotes the total number of observation items whose definitions are maintained. Statistical relationships among items are examined according to the research purpose. Relationships among the layers form an explanatory model organizing practical observations and related research; causal relationships under particular conditions are evaluated in specific studies. Research on changes in

perception and behavior accompanying changes in spatial structure continues with records of product, people, temperature, time, information, and other conditions, including alternative explanations.

5 A connection model based on observation records

5.1 Connection principles

The unit connecting the systems is an observation record retaining its conditions and evidence [10,12]. For example, ICES texture findings inform the SEAS state axis, while conformity to product specifications and conformity to provision goals are judged independently. TDC cross-section discrimination records can also inform the SEAS structure axis. Independence here means separation of judgment references and procedures. Statistical relationships between scores are treated as a separate object of analysis.

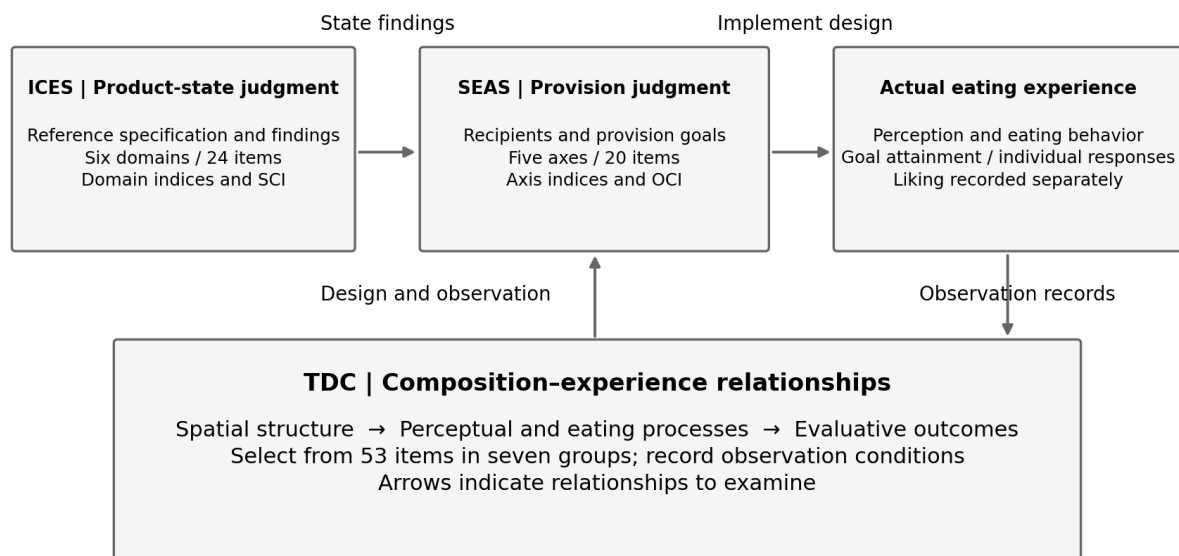
Figure 1 shows a connection model identifying the roles of the three systems. Product specifications, provision goals, and individual responses are retained as distinct references or records; observations after

design changes inform subsequent judgments. Findings are shared while each system's overall index is displayed separately with its specific meaning retained.

5.2 Shared records and independent judgments

To connect observations with judgments, common records retain sample identifiers, condition identifiers, observation times, observers, item identifiers qualified by system name, units or scales, findings, reference specifications or purposes, judges, and the version of the criteria used [10,12]. Identifiers, abbreviated as IDs in Figure 1, are codes linking records for the same samples, conditions, and items. When temperature or surface conditions change after provision, the changed state is recorded through observation at that time.

To preserve evaluation independence, brand information and others' ratings that are unnecessary for judging product state are masked as appropriate to the purpose. Professional provision judgments are recorded, where possible, before the eater's final response is known. When judgments are revised after consulting responses, the original record and reason for change are retained. This allows design-stage predictions to be compared with actual outcomes.



Connect through common sample IDs, condition IDs, and recording times; retain distinct indices.

Figure 1 Connection model for the three systems

Boxes identify objects of recording or judgment; arrows indicate reference to findings, implementation, or observation. Arrows within TDC represent conceptual relationships organizing composition, processes, and outcomes. Causal relationships are evaluated in specific comparative studies.

5.3 Conditions of application and comparative research

The model directly addresses cheese comparative learning and provision design where samples, provision goals, and observation conditions can be identified. Records can be connected when the same lot or corresponding samples, a time window within which state findings remain applicable, and the relevant criteria version can be established. SCI values based on different product references or OCI values based on different provision goals are not used to create a common ranking.

Product-state conformity, attainment of provision goals, and individual liking are judged using their respective evidence. Effects of composition are examined in relation to recipients, viewpoints, skills, and provision methods. This paper compares a model

explaining outcomes through state findings with a model that adds provision judgments. The usefulness of the connection model is assessed through changes in predictive performance and traceability of judgment evidence, recording improvements, equivalent outcomes, and decreases in performance.

6 Operational rules for judgments and indices

6.1 Four-level judgments and inability to judge

ICES and SEAS use valid judgments from 1 to 4 and a code of 0 when a judgment cannot be established. The statements defining the levels differ between the systems. Zero is not the lowest score or a fifth scale category; it indicates insufficient information, samples, conditions, or other prerequisites. Its reason is recorded, and it is excluded from both the numerator and denominator of the index [1,11,12].

Table 2 Judgment statements and treatment in records

Code	SEAS	ICES	Treatment
4	Conforming	Conforming	Valid judgment
3	Broadly conforming	Minor departure	Valid judgment
2	Borderline	Major departure	Valid judgment
1	Nonconforming	Nonconforming	Valid judgment
0	Unable to judge / not established	Unable to judge / not established	Record the reason; exclude from aggregation

TDC includes a supplementary five-level assessment, from 0 to 4, recording the degree to which an observed item meets its criteria [10]. Zero is a valid assessment meaning that the item was observed but did not meet basic criteria. Not observed or unable to judge, and non-applicability specified in advance, are recorded separately from numerical assessments. These numbers and the four-level SEAS judgments have different meanings and are judged using their respective criteria. The supplementary assessment is used for items selected according to purpose.

6.2 Conditions for calculation

An axis or domain index is calculated when at least three of its four items have valid judgments. Here j denotes an axis or domain, i an item, V_j the set of valid items in that axis or domain, n_j the number of valid items, x_i the score for item i , and I_j the axis or domain index. The symbol Σ denotes summation over the

specified items. Equation 1 gives the calculation rule [1,11,12].

$$I_j = \frac{100}{3n_j} \sum_{i \in V_j} (x_i - 1) \quad (1)$$

For SEAS, the provision conformity index OCI is displayed when all five axes are established, using an equally weighted mean of the axis indices. For ICES, the state conformity index SCI is displayed analogously when all six domains are established. Equation 1 is a linear transformation: subtract the lower endpoint, 1, from the mean of valid scores, divide by the range width, 3, and multiply by 100. Because that mean lies between 1 and 4, the index lies between 0 and 100, with 1 mapping to 0 and 4 to 100. This transformation defines the display range; equal spacing of the scale is treated as an operational assumption in Section 6.3. The subscripts of I in Equations 2 and 3 are the axis and domain codes defined in Section 4; for example,

I_{ST} denotes the state-axis index and I_{EX} the external-state domain index.

$$OCI = \frac{I_O + I_{ST} + I_C + I_S + I_M}{5} \quad (2)$$

$$SCI = \frac{I_{EX} + I_{IN} + I_{TX} + I_{AR} + I_{FL} + I_{MA}}{6} \quad (3)$$

The indices are operational summaries on a 0–100 range, not probabilities, quality percentages, liking scores, or pass/fail decisions. If all scores are 1, the index is 0, a result based on valid judgments that differs from the 0 code for inability to judge. If any required axis or domain is not established, the overall index is not calculated. Intermediate values are not rounded; displayed values are rounded to one decimal place [1].

6.3 Aggregation assumptions and comparability

Equal spacing of the four levels and equal weighting of axes or domains are operational assumptions of the calculation rules in this paper. Measurement intervals and the relative importance of items are evaluated in specific studies. Comparisons report the original item judgments, findings, number of valid items, and items that could not be judged alongside the indices.

For example, hypothetical scores of 4, 3, 3, and 2 produce an index of 66.7, while scores of 4, 3, 3, and 0, with the last item unable to be judged, produce 77.8. The difference reflects the composition of retained items and is interpreted separately from changes in product state or provision. Research should prioritize designs observing the same items and report the reasons and frequency of missing judgments. When indices calculated from different valid-item sets are compared, distributions for commonly observed items can be examined as supplementary information, distinguishing the official indices from supplementary analyses.

Averages can also obscure important departures in individual items. Practice therefore retains item-level records and identifies appropriate responses individually. Safety is checked through separate information and management procedures. Conversion between SCI and OCI, or conversion to the 20-point score of the Japan National Cheese Evaluation, is not included in the operation of this framework [1].

6.4 Individual and panel-representative judgments

For ICES and SEAS, the operational rules require an official panel of at least three assessors, retaining

individual judgments and calculating indices from the median judgment for each item [11,12]. Averaging assessor-specific OCI or SCI values and calculating an index from item medians involve different calculation orders and generally give different results. Reports identify whether judgments are individual or panel-representative.

For research, handling of item medians with an even number of valid assessors, the minimum number of valid assessors, and cases in which all assessors are unable to judge are specified before the study and stated in reporting. Rounding representative values and imputing missing judgments also follow prespecified rules. The requirement of at least three of four items in Section 6.2 concerns valid items within an axis or domain and is managed separately from assessor-number requirements.

7 Hypothetical application of the connection model

7.1 A learning task comparing cross-sections

A learning task comparing the cross-sections of two cheeses is presented as a hypothetical example constructed to explain the method. The intended learners are encountering these cheeses for the first time, and the purpose is to identify differences between the cross-sections and explain them from observation.

First, the product-specific reference specifications and evaluation times are recorded and product state is checked using ICES. Where product references differ, SCI values are not interpreted as a common quality ranking. Two presentation conditions are then planned using the same products, slice dimensions, quantities, temperatures, observation times, vessels, and lighting, varying only slice orientation. Changing the cut at the same time could change the proportions of rind and interior and make arrangement effects difficult to distinguish from sample-composition effects; this example therefore varies orientation alone.

TDC items selected are direction within the plane, discrimination of cross-sections or layer boundaries, and attainment of the provision goal. Correct discrimination, response time, and explanations are recorded using their respective procedures. SEAS uses these records as evidence for the structure-axis judgment while also judging the other four axes using their specified items. Individual liking is measured

separately, and its relationship with ease of discrimination is examined individually.

7.2 Identifying improvements from combinations of outcomes

Table 3 presents hypothetical combinations of outcomes constructed to explain the method.

Table 3 Hypothetical outcome combinations and subsequent checks

Combination of observations or judgments	Subsequent checks
Product state conforms, but cross-sections are difficult to discriminate	Check orientation, viewpoint, lighting, and explanation procedure separately
Discrimination improves, but liking is unchanged	Record learning-goal attainment and preference as separate outcomes
Provision judgments are high, but the goal is not attained	Review judgment evidence and optimal ranges against individual behavior
Texture and aroma change over time	Check presentation time and temperature history to establish comparable states

The value of this procedure is that it records which observations support which judgments and which conditions to examine next. Building on knowledge accumulated through practice, research continues through comparison with existing recording methods to clarify usefulness and conditions of application.

8 Testable propositions

This section presents four comparative propositions for examining the usefulness and applicability of the practice-derived framework. The propositions are specified for subsequent comparative research; the contribution here is their formulation and the definition of evaluation procedures. Comparison conditions and evaluation items are specified to assess the degree of support for each proposition.

8.1 Proposition 1 Separating judgment objects and tracing explanations

A method that separates state findings, provision judgments, and individual experience and links them through records with specified conditions is predicted to help third parties identify judgment evidence and targets for improvement. Three conditions are compared: free-text records; structured records with the same fields but without links among records; and connected records linked through evidence IDs. Observation information, recording time, and experience levels are kept comparable. Evidence

Recording the systems separately specifies whether to examine the product, provision procedure, learning task, or preference. Causes are investigated by reassessment with selected conditions changed.

identification rate, misidentification of judgment objects, and checking time are measured. Connected versus structured records is the primary comparison for the usefulness of the connection itself.

Support for the proposition is assessed jointly through evidence identification and recording burden. Equivalent or lower identification rates, and reductions in important recorded observations associated with increased burden, are also documented to examine conditions of application. Evaluators assess record content with the method's name and developer masked.

8.2 Proposition 2 Shared reference specifications and agreement

For evaluation of the same samples, advance sharing of reference specifications, item definitions, and judgment examples is predicted to reduce disagreement in ICES item judgments compared with reliance on each assessor's implicit references. Training exposure and sample conditions are matched, and independent evaluation records are retained. Convergence on similar scores is distinguished from the ability to discriminate relevant differences in state.

Agreement after references are shared informs examination of the clarity of definitions and examples, overlap among items, and specificity of reference specifications. Judgments based on different reference

specifications are interpreted in relation to their respective specifications and evidence.

8.3 Proposition 3 Manipulation of composition and perception and behavior

In a cross-section comparison task with matched product state and presentation conditions, orienting the target cross-section toward the observer is predicted to improve discrimination accuracy or judgment speed relative to an arrangement that makes it less visible. TDC spatial structure serves as the manipulated condition, cross-section discrimination as a process record, and task correctness as an outcome record. Comparison order, learning through repetition, and differences in visual acuity or prior knowledge are considered.

The prediction concerns discrimination accuracy and judgment speed in the defined task and population. Taste intensity, liking, and purchase intention are measured as outcomes for their respective research questions. Studies explaining the relationship between perceptual discrimination and task correctness through processes address temporal order and alternative explanations when examining mediated causal relationships.

8.4 Proposition 4 Provision judgments and independent goal attainment

In settings with recipients and purposes specified in advance, SEAS provision judgments fixed without knowledge of the main trial's outcomes are predicted to add predictive information beyond ICES state judgments for subsequently and independently measured goal attainment. Standardized preliminary provision using separate samples or participants can provide judgment evidence. In comparative learning, task correctness or explanation accuracy serves as the outcome; main-trial participant responses are not included in the advance SEAS judgment.

Judgments and conditions obtained in preliminary provision are fixed before the main trial and examined with independent recipients or settings. Items lacking the evidence required before implementation receive the 0 code. Where the specified conditions are not met, OCI is not generated; usable items or axes are designated in advance for analysis. Models using baseline information, then adding ICES findings, then SEAS findings are compared. Using the same outcome both for SEAS judgment and goal attainment is avoided, and liking is treated as a separate outcome.

9 Development into specific empirical studies

9.1 Examining definitions and content

When designing specific studies, each item is mapped to its observation object, definition, reference, respondent, time, scale, and judgment examples. Review from the perspectives of production, maturation, sales and provision, sensory evaluation, and consumer understanding examines relevance, shared interpretation, overlap, and coverage. This examination of item content is combined with assessment-performance studies using observation data to clarify evidence for validity.

A TDC item may serve as a manipulated condition in one study and an observed process or outcome in another, so its role is specified before research begins. For SEAS and ICES, models in which multiple components constitute overall conformity are distinguished from models in which items reflect a common latent factor. Use of factor analysis or internal consistency is guided by a measurement model appropriate to item construction and research purpose.

9.2 Repetition and independent evaluation

Repeated presentation of samples under matched conditions then examines item response distributions, frequency of inability to judge, within-assessor repeatability, and agreement among assessors. Presentation order and sample codes are controlled, and independent records made before discussion are retained. Numbers of assessors, samples, and repetitions are determined by the planned analysis and required precision.

Agreement measures or models appropriate to the ordered four-level judgments are considered. Reports show which items disagree and in which direction, alongside coefficients. When chance-corrected agreement measures are used, the effect of concentration in particular response categories is also examined. Item-level analysis is primary, and assessment performance is examined from multiple perspectives alongside OCI and SCI.

9.3 Comparative studies of composition and provision

Comparisons of composition specify the factors manipulated at a given time and use appropriately matched samples from the same product. Order is balanced where repeated presentation is suitable;

between-participant allocation is also considered for tasks with strong learning or memory carryover. Temperature, quantity, time, information, and viewpoint are recorded. Changes in exposed surface area or aroma release accompanying a manipulation are distinguished as conditions to hold constant or processes to investigate.

Primary outcomes, exclusion criteria, missing-data treatment, and analytical methods are specified before results are examined. Analysis of condition effects accounts for clustering of responses from the same participants, samples, or settings. Explanations selected during exploration are examined using new samples and participants. The direction and magnitude of effects, together with conditions showing equivalent outcomes, are recorded to clarify applicability.

9.4 Usefulness in education and practice

Practical usefulness is evaluated through recording time, understanding of explanations, specificity of improvement methods, and outcomes after changes, as well as rating agreement. Feedback to producers develops hypotheses about causes by comparing sensory findings with production, maturation, and storage records. Learners are asked to distinguish sensations, reference-based judgments, and matters to examine next.

Studies using data follow the procedures of the implementing institution for appropriate ethics

review, participant consent, and personal-information management. External evaluators are included, and influences of developer expectations and training content on outcomes are examined. Research records and formal evaluation or certification records are managed with their respective purposes and conditions specified.

10 Discussion

10.1 Scholarly contribution

The contribution is to formalize product state, provision conformity, and eating experience as a research framework retaining distinct objects and references of judgment, and to develop propositions for comparing practical usefulness. It builds on prior research addressing sensory vocabulary, specification conformity, meal context, and the distinction between expectations and experience, together with the textbook and operational documents identified in Section 3.1.

This structure provides a method for describing product-state conformity, attainment of provision goals, individual preferences, learning outcomes, and flavor evaluation in relation to their respective judgment objects and conditions. For example, changes in learning outcomes and flavor evaluation following a composition change can be recorded separately and their relationship examined. Additional explanatory value and recording usefulness are evaluated using the propositions in Section 8.

Table 4 Contribution defined through comparison with prior research

Comparison	Shared foundation	What this paper specifies
Cheese vocabulary [3,4]	Shared attributes through terms and reference examples	Procedures comparing the same observations separately with product specifications and provision goals
ISO 22935-3 [13]	Scoring dairy-product sensory properties against specifications	Conditions for referring ICES findings to SEAS without converting indices
FAMM [14]	Meal services encompassing food, people, and environment	Records tracing cheese-state diagnosis, purpose-specific provision, and independent experiential outcomes
Grunert's quality-perception framework [15]	Distinction between expected and experienced quality	Product specifications and provision goals retained as separate judgment references alongside consumer expectations

The right column identifies this paper's conceptual positioning through comparison with the cited research. Comparative practical performance is

addressed in studies corresponding to the propositions in Section 8.

The paper presents a connection model that makes judgment evidence traceable, the meaning and application conditions of each scale, and four propositions that can be compared with existing approaches. Proposition 1 uses matched recording fields to distinguish effects of structured recording in general from added effects of connecting the systems. Proposition 4 examines what provision judgments add about independent outcomes once product state is known.

10.2 From educational practice to public reference

The framework emerged from organizing knowledge of observation, explanation, and composition cultivated in the author's educational and provision practice into textbooks and criteria systems. Education shares sensory descriptions and judgment reasons, while provision practice examines composition and methods in relation to recipients and purposes. Publishing this practical foundation as a document with definitions, sources, and operational rules enables learners, educators, and practitioners to refer to and examine the same descriptions. Applications to producer support and dialogue with consumers can also be developed with defined recipients and conditions. The magnitude of outcomes in each setting is evaluated separately from this paper's conceptual analysis.

Continued use in society requires a balance between recording detail and practical burden. Research retains detailed conditions, while practice uses displays and references appropriate to purpose. Trial recording forms with fewer items are distinguished from complete evaluation and indices using all items as described in Sections 4–6. Cases of inability to judge and disagreement encountered in practice inform revision of definitions, teaching materials, and judgment examples.

10.3 Scope of evidence and continuing research

The contribution established here is the conceptual formalization of a practice-derived framework. Research will continue through analysis of practice records and comparative studies examining effect sizes under specific conditions, agreement among assessors, and applicability to other recipients and settings. Methods for aggregating judgments from multiple assessors are also examined with explicit purposes and conditions. External literature was selected for this paper's purpose, and its scholarly positioning rests on the literature compared here.

Examination of item content, agreement, and repeatability will further clarify the purposes and conditions under which the item counts, four levels, and minimum of three valid items function. Equal weighting is the operational choice set out in Section 6.2, distinct from empirical optimality. Differences among products, languages, training histories, cultures, and provision settings are also research objects for examining applicability. Research develops through external examination by researchers and practitioners and testing under independent conditions, considering the influence of the developer's selection and interpretation on the organization of the framework.

11 Conclusion

Drawing on twelve years of cheese education practice and relevant research, this paper presented an evaluation methodology connecting ICES, SEAS, and TDC. It organized product-state conformity judgments, provision-conformity judgments relative to purpose, and composition–experience relationships while retaining the evidence for each system's judgments. Observation records identifying samples, conditions, times, and items serve as the unit of connection, making the conceptual model and operational principles explicit.

This paper makes an evaluation methodology grounded in vocational education and provision practice available as a research paper for external citation and examination. Practice and research will continue, clarifying evidence for specific effects, assessment performance, and applicability through analysis of existing records and comparative studies appropriate to the questions addressed. The four propositions in Section 8 provide guidance for developing that research.

Research declarations

Data and materials availability. This is a document-based study of concepts and operational structure. The foundational textbook, theoretical source, and specifications are identified in references [1,10–12]; external research and standards are identified by their corresponding reference numbers. Principal definitions, objects of judgment, operational rules, and formulas are provided in Sections 4–6, together with an explanatory figure, a hypothetical example, and comparative propositions. Analyses using participant responses are addressed in separate

practice-based studies specifying the periods covered, units of aggregation, and available records.

Author's relationship to the activities studied. The author chairs CAFAJ and is involved in developing the systems and textbooks examined and in the

educational business. Related textbooks are sold to learners. The framework's organization is grounded in the author's own practice; this relationship is disclosed to support external examination and subsequent research.

References

- [1] Kaneko T. KR Knowledge Reference: Common-Foundation Textbook for Fromager Certification, Upper Volume: Cheese Culture, Science and Evaluation. CAFAJ Research Publishing; 2026.
- [2] Jaakkola E. Designing conceptual articles: four approaches. *AMS Review*. 2020;10:18–26.
- [3] Drake MA, McIngvale SC, Gerard PD, Cadwallader KR, Civile GV. Development of a Descriptive Language for Cheddar Cheese. *Journal of Food Science*. 2001;66(9):1422–1427.
- [4] Drake MA, Gerard PD, Wright S, Cadwallader KR, Civile GV. Cross validation of a sensory language for Cheddar cheese. *Journal of Sensory Studies*. 2002;17(3):215–227.
- [5] Michel C, Velasco C, Gatti E, Spence C. A taste of Kandinsky: assessing the influence of the artistic visual presentation of food on the dining experience. *Flavour*. 2014;3:7.
- [6] Harrar V, Spence C. The taste of cutlery: how the taste of food is affected by the weight, size, shape, and colour of the cutlery used to eat it. *Flavour*. 2013;2:21.
- [7] International Organization for Standardization. ISO 6658:2017 Sensory analysis — Methodology — General guidance. Geneva: ISO; 2017.
- [8] International Organization for Standardization. ISO 8586:2023 Sensory analysis — Selection and training of sensory assessors. Geneva: ISO; 2023.
- [9] International Organization for Standardization. ISO 11132:2021 Sensory analysis — Methodology — Guidelines for the measurement of the performance of a quantitative descriptive sensory panel. Geneva: ISO; 2021.
- [10] CAFAJ. TDC Integrated Theoretical Master Document [theoretical material]. 2026.
- [11] CAFAJ. SEAS KR Formal Specification: 官能適用評価基準 [specification]. 2026.
- [12] CAFAJ. ICES KR Formal Specification: International Cheese Evaluation Standard [specification]. 2026.
- [13] International Organization for Standardization. ISO 22935-3:2023 Milk and milk products — Sensory analysis — Part 3: Method for evaluation of compliance with product specifications for sensory properties by scoring. Geneva: ISO; 2023.
- [14] Gustafsson I-B, Öström Å, Johansson J, Mossberg L. The Five Aspects Meal Model: a tool for developing meal services in restaurants. *Journal of Foodservice*. 2006;17(2):84–93.
- [15] Grunert KG. Current issues in the understanding of consumer food choice. *Trends in Food Science & Technology*. 2002;13(8):275–285.

Scope of source access. References [2,5,6] were consulted through publishers' public full texts; [3,4,15] through institutional abstracts and bibliographic records; [14] through the publisher's abstract and bibliographic record; and [7,8,9,13] through official bibliographic records and public summaries of the International Organization for Standardization. Discussion of standards concerns their scope and methodological positioning. English titles of Japanese source materials are descriptive translations.