EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Developing and Validating MCA-Equivalent Summative Assessment in Combinatorics Using Aiken's V Method

Nurjihan Nabilah[1], Sovy Dia Permata Sari[1], Yurizka Melia Sari[1, 2, a)]

[1]*Universitas Negeri Surabaya*
[2] *EduSTEM Centre Universitas Negeri Surabaya*
Ketintang Street, Ketintang, Gayungan, Surabaya, East Java, Indonesia, 60231

[a)] nurjihan.23037@mhs.unesa.ac.id

Abstract. This study aims to analyze the content validity of summative assessment items in Combinatorics that are aligned with the Minimum Competency Assessment (MCA) framework. The study was motivated by the need for valid and high-quality assessment tools that support meaningful and sustainable mathematics learning. This study employed a research and development (R&D) design but it focuses primarily on the Develop stage, which aims to obtain a valid summative assessment instrument. The Aiken's V method was used to measure the content validity of nine multiple-choice items developed according to the MCA numeracy domains, cognitive levels, and contexts. Three mathematics education experts served as validators. The analysis showed that the Aiken's V indices ranged from 0.64 to 0.73, indicating medium validity. These results suggest that the instrument can be used after minor revisions. The validated instrument provides a reference for educators to design MCA-oriented summative assessments that enhance students' mathematical literacy and reasoning skills. Future research may apply other validation methods such as Gregory's analysis to strengthen the evidence of validity.

Keywords: Aiken's V Method; Combinatorics; Content Validity; Minimum Competency Assessment; Mathematical Literacy

Abstrak. Penelitian ini bertujuan untuk menganalisis validitas isi instrumen asesmen sumatif pada materi Kombinatorika yang disusun setara dengan kerangka Asesmen Kompetensi Minimum (AKM). Kebutuhan terhadap instrumen asesmen yang valid dan berkualitas tinggi menjadi dasar dalam mendukung pembelajaran matematika yang bermakna dan berkelanjutan. Penelitian ini menggunakan desain penelitian dan pengembangan (R&D) tetapi berfokus terutama pada tahap Pengembangan, yang bertujuan untuk memperoleh instrumen penilaian sumatif yang valid. Metode Aiken's V digunakan untuk mengukur validitas isi terhadap sembilan butir soal pilihan ganda yang dikembangkan berdasarkan domain, konteks, dan tingkat kognitif literasi numerasi AKM. Validasi dilakukan oleh tiga ahli pendidikan matematika. Hasil analisis menunjukkan bahwa nilai indeks Aiken berkisar antara 0,64 hingga 0,73 yang menunjukkan tingkat validitas sedang. Hal ini menunjukkan bahwa instrumen asesmen sumatif tersebut dapat digunakan setelah dilakukan revisi kecil. Instrumen yang telah divalidasi ini diharapkan dapat menjadi acuan bagi pendidik dalam merancang asesmen sumatif berbasis AKM yang mendorong pengembangan kemampuan literasi numerasi dan penalaran matematis siswa. Penelitian lanjutan disarankan menggunakan metode validitas lain seperti metode Gregory untuk memperkuat bukti validitas.

Kata kunci: Asesmen Kompetensi Minimum; Kombinatorika; Literasi Numerasi; Metode Aiken's V; Validitas Isi

Available online at journal homepage:
http://ejournal.iainkerinci.ac.id/index.php/edumatika
Email: edumatika@iainkerinci.ac.id
DOI: https://doi.org/10.32939/ejrpm.v8i2.5222

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

INTRODUCTION

Summative assessment is one of the crucial components in the mathematics learning assessment process. The assessment should reflect comprehensively the achievement of students' mathematics learning outcomes in accordance with the learning objectives. In line with this, summative assessments summarize student learning outcomes with the aim of gaining a holistic view of student performance (Chytry & Kubiatko, 2021). In mathematics education, such assessments are typically conducted at the end of a unit or topic to evaluate the effectiveness of learning and measure students' conceptual understanding. 21st century mathematics learning emphasizes the development of critical thinking skills and the ability to link knowledge with real-world contexts, so that students not only master concepts, but are also able to apply them relevantly in everyday life (Lestari & Ali, 2023). In this regard, effective assessment is central to developing 21st-century competencies such as problem solving, creativity, and reasoning (Griffin et al., 2012). These assessments often form the basis for decisions that teachers make both regarding further teaching strategies and individualized assistance for learners (Hilden et al., 2022). Therefore, a well-designed summative assessment plays an essential role in guiding teachers' instructional decisions and providing meaningful feedback for students.

To ensure that a summative assessment accurately measures what it intends to measure, content validity becomes a key requirement. Lawshe (1975) emphasized that expert judgment plays a central role in evaluating the essentiality of each item within an instrument. Content validity refers to which the assessment is able to measure what it is supposed to measure (Aydinyer & Ubuz, 2025). This process usually involves expert judgment from validators who assess the relevance and representativeness of each item. Expert judgment is required, which usually involves three peers as validators (Pandra & Aswarliansyah, 2023). The results of these expert judgments are then analyzed using statistical techniques such as Aiken's V coefficient **to** determine the level of validity (Zakiyah & Kartika, 2024). Consequently, the careful selection of validators in accordance with the research field is crucial to ensure accurate and relevant assessments, especially for complex mathematical topics.

One of mathematics topics that poses difficulties in creating assessment tools because of its intricacy is Combinatorics. Combinatorics, as noted by Lockwood (2013), involves understanding both the structure and reasoning processes used to count and arrange elements systematically. This topic possesses distinct features, as resolving combinatorial problems necessitates strong analytical thinking abilities (M. R. Putri et al., 2020). In the learning process, students are not only required to memorize but also to develop logical strategies in solving problems. Research by Batanero et al. (1997) also highlights that students commonly struggle with combinatorial reasoning due to its abstract nature and the need for flexible thinking. Hence, a meaningful and well-validated assessment

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

is necessary to evaluate students' conceptual and procedural understanding effectively. In line with the findings that concept understanding will be well embedded, if the learning process and assessment are meaningful (Lukman & Setiani, 2018).

In the effort to enhance the quality of assessments, developing items aligned with the Minimum Competency Assessment (MCA) framework has become increasingly important. MCA is a tool to measure students' cognitive abilities based on reading literacy and numeracy literacy (Siahaan et al., 2022). According to OECD (2021), numeracy literacy refers to the ability to apply mathematical reasoning in a variety of real-life contexts, forming one of the key dimensions assessed in MCA. MCA is implemented in the Indonesian education system with the aim of improving the quality of learning and student learning outcomes. Therefore, the main focus of MCA is the fulfillment of students' reading literacy and numeracy literacy skills (Andrianti & Rahayu, 2022). Assessment items grounded in numeracy literacy should integrate contextual reasoning and problem-solving, reflecting real-world mathematical practices (Goos et al., 2014). In addition, the use of MCA questions can also evaluate mathematical logical intelligence, where students with high mathematical logical intelligence are able to interpret the steps of solving the question until they find the correct answer (Andrianti & Rahayu, 2022). Thus, MCA-oriented assessments are particularly suitable for Combinatorics topics that require systematic and logical approaches.

Several studies have analyzed the content validity using the Aiken's V method in various contexts (Y. F. Putri et al., 2022), such as validating mathematics textbooks (Prafianti et al., 2023), problem-solving skills instruments (Kania et al., 2024), the development and validation of mathematical thinking instruments (Alhunaini et al., 2020), the validity of formative assessments (Nurjanah et al., 2023) and the validity of summative assessments (Pandra & Aswarliansyah, 2023), as well as the development of valid and reliable summative tests in trigonometry (Tandas, 2021). However, limited research has focused specifically on the content validity of MCA-equivalent summative assessments for Combinatorics. Meanwhile, studies show that students often experience difficulties in solving Combinatorics problems, especially in cases of permutations and combinations (Pratiwi et al., 2023). These findings suggest that students' understanding of Combinatorics concepts remains incomplete, and existing assessments may not yet fully capture the unique characteristics of this topic.

Based on these gaps, this study aims to analyze the content validity of MCA-equivalent summative assessment items in Combinatorics using Aiken's V method. Previous studies have not extensively discussed MCA-equivalent summative assessment instruments for Combinatorics. The study seeks to contribute to the development of valid assessment instruments that can support integrated and meaningful learning in Combinatorics. The expected outcome is an assessment tool

EDUMATIKA: Jurnal Riset Pendidikan Matematika *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

with strong content validity that can serve as a reference for teachers in designing summative assessments aligned with learning objectives and numeracy literacy competencies.

METHOD

This study employed a research and development (R&D) design, which aims to produce certain products and test the effectiveness of these products (Prafianti et al., 2023). The product developed in this study is a summative assessment instrument for Combinatorics material designed to be equivalent to the numeracy component of the Minimum Competency Assessment (MCA). This research model uses Thiagarajan's 4D model, which consists of four sequential stages, namely define, design, develop, and disseminate (Ramadani et al., 2024). Figure 1 illustrates the flow of these stages.

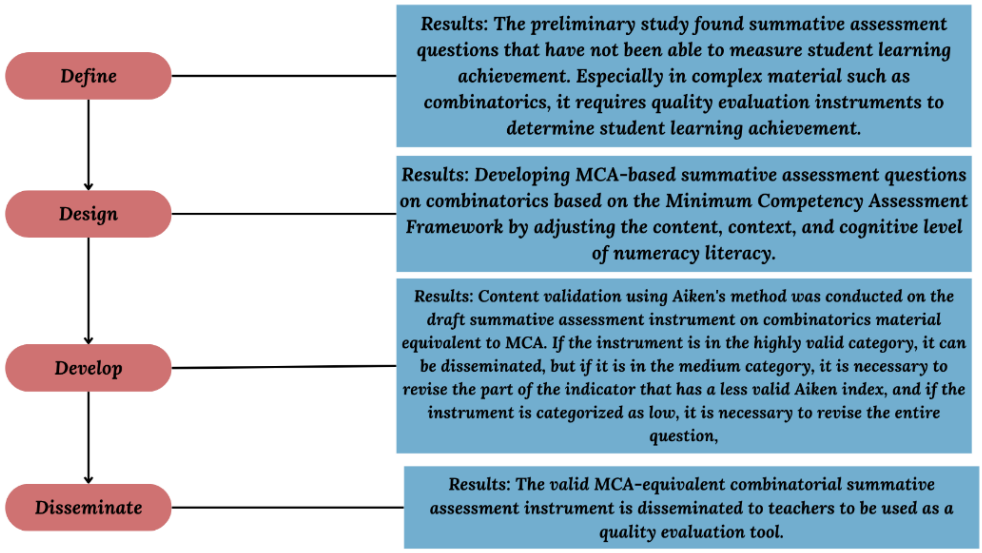

Figure 1. Research Stages

At the define stage, researchers identified problems related to existing summative assessment items on Combinatorics material that did not effectively measure learning objectives or support numeracy literacy. This stage included an analysis of curriculum standards, learning outcomes, and characteristics of MCA questions. The findings from this phase served as the foundation for developing new assessment items aligned with MCA standards and Combinatorics learning goals.

Furthermore, the design stage focused on constructing blueprints for summative assessment items aligned with MCA numeracy contexts. The design process referred explicitly to the Minimum Competency Assessment Framework, which maps each item according to domain, context, and cognitive level. Researchers developed item specifications considering the Combinatorics competencies in high school, ensuring the content's relevance to both the curriculum and numeracy literacy domains. This stage resulted in the preparation of test specifications and initial drafts of assessment items.

EDUMATIKA: Jurnal Riset Pendidikan Matematika *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Meanwhile, at the develop stage, the summative assessment items were constructed and refined according to the design specifications. The instrument focused on the Data and Uncertainty domain within a personal context familiar to students. The items were structured at three cognitive levels of numeracy: Knowing, Applying, and Reasoning.

This study focuses primarily on the Develop stage, which aims to obtain a valid summative assessment instrument that can serve as an evaluation tool for Combinatorics learning. Although the 4D model includes the Disseminate stage, this paper reports only up to the validation process. The develop stage is considered crucial as it determines the quality and accuracy of the assessment instrument. To ensure the rigor of the validation process, this study followed the principles of expert-based validation as proposed by Polit and Beck (2006), who compared several approaches such as Content Validity Ratio (CVR) and Content Validity Index (CVI) before recommending Aiken's V as a practical alternative for educational contexts. Furthermore, Taherdoost (2016) emphasized that methodological consistency and expert selection are critical factors determining the trustworthiness of instrument validation.

The draft instrument consisted of nine multiple-choice items: three at the Knowing level, three at the Applying level, and three at the Reasoning level. Each item was written at a higher level of difficulty to reflect the analytical nature of MCA questions (Prafianti et al., 2023). Table 1 presents an excerpt of one of the items developed.

Table 1. Excerpt of Summative Assessment Instrument for MCA Equivalent Combinatorics Materials

Level	6 (Grades 11 and 12)
Domain	Data and Uncertainty
Subdomain	Conditional probability and independent events
Context	Personal
Competency	Differentiate conditional probability events from independent events
Cognitive Level	Knowing
Problem Form	Multiple Choice

Andi throws one red and one white dice simultaneously. On the first roll, the number that appears is the number 2 on the red die and the number 1 on the white die. On the second throw, the number that appears is the number 4 on the red dice and the number 2 on the white dice. Then on the third roll, the number that appears is the number 6 on the red die and the number 5 on the white die. Are these events conditional probabilities or mutually exclusive events?

A. The events are mutually exclusive, as the second and third throws are influenced by the first.
B. The events are mutually exclusive, as the second throw is always larger than the first.
C. Conditional odds, as the first toss is smaller than the second and third.
D. Events are mutually exclusive, as the first throw does not affect the second and third throws.
E. It cannot be determined that the event is a conditional probability or a mutually exclusive event from the information provided.

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

The validity test in this study focused on content validity. Validation was conducted by three experts in mathematics education who served as validators. They were asked to assess the relevance, clarity, and alignment of the assessment items with learning objectives and numeracy literacy standards. The validation sheet consisted of 17 indicators, which covered three main aspects: content suitability, item construction, and language clarity. Examples of the indicators are presented in Table 2.

Table 2. Indicator on the Validation Sheet

Indicator	Items Examples
Indicator 1	Questions are in accordance with the learning objectives to be achieved
Indicator 2	Questions are in accordance with the competencies measured
Indicator 3	Answer choices are homogeneous and logical
Indicator 4	There is only one correct answer key
Indicator 5	Questions are in accordance with the cognitive domain measured

Each indicator was rated using a Likert scale from 1 to 5, where 1 = Very Irrelevant and 5 = Very Relevant (Table 3). The instrument will be valid if the experts believe that the instrument measures the things that will be measured (Pandra & Aswarliansyah, 2023).

Table 3. Score Categories on the Validation Sheet

Scale	Category
1	Very Irrelevant
2	Not Relevant
3	Moderately Relevant
4	Relevant
5	Very Relevant

Quantitative analysis of the validation results was carried out using Aiken's V index (Aiken, 1980; Aiken, 1985). The Aiken's V coefficient was calculated using the following formula:

$$V = \frac{\sum s}{n(c-1)} \text{ where } \mathrm{s} = \mathrm{r} - L_0$$

r = rating given by a validator,
L_0 = lowest possible score in the scale,
n = number of validators, and
c = number of categories on the scale.

All calculations were performed using Microsoft Excel to determine the Aiken's V index for each item. The interpretation of validity levels followed Susantini et al. (2025) as shown in Table 4.

Table 4. Validity Categories

Aiken Index	Validity Level	Interpretation
$V \leq 0{,}4$	Less Valid	Requires full revision of the item
$0{,}4 \leq V \leq 0{,}8$	Moderately Valid	Requires partial revision of weak indicators
$0{,}8 < V$	Highly Valid	Item can be used without revision

EDUMATIKA: Jurnal Riset Pendidikan Matematika *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Based on Table 4, an item is considered acceptable if it achieves an Aiken index of $V \geq 0.80$. Items with indices between 0.40 and 0.80 require revision, while those below 0.40 are deemed invalid and must be completely revised.

RESULTS AND DISCUSSION

MCA-Equivalent Summative Assessment in Combinatorics

The developed product is a summative assessment instrument for Combinatorics material equivalent to the Minimum Competency Assessment (MCA). It consists of nine multiple-choice items, representing the three cognitive levels of numeracy: Knowing, Applying, and Reasoning. Specifically, three items assess the Knowing level, three the Applying level, and three the Reasoning level. In addition, two items employ socio-cultural contexts and seven use personal contexts in alignment with MCA characteristics. Tables 5–7 present examples of assessment items developed for each cognitive level.

Table 5. Example Item 1 (Knowing Level)

Level	6 (Grade 11 and 12)
Domain	Data and Uncertainty
Subdomain	Conditional probability and unconditional probability
Context	Personal
Competency	Identify the difference between conditional odds and unconditional odds
Cognitive Level	Kowing
Problem Form	Multiple Choice

Take a look at the two statements below!

Statement 1:

Naomi is an only child. She has many cousins because Naomi's mother is the youngest of two siblings while Naomi's father is the eldest of three siblings. Which, all the siblings of his parents already have children. So when she was on vacation in London, she intended to give random gifts to her cousins. She gathered them together in a big box. Naomi bought 6 toy cars and 6 rubik toys. She predicted that the first randomly drawn gift would be a toy toy toy. So Naomi states that the probability of the next two prizes being drawn at random should be calculated based on the remaining prizes.

Statement 2:

SMAN 2 Surabaya will hold an election for student council president. This student council can be a place for students who want to develop their skills. Especially being the head of the student council. When becoming the student council chairman, the skills they can develop are public speaking, time management, responsibility, coordination with fellow friends, and so on. When the registration was opened, there were 5 students who registered as candidates for student council chairman. They had to go through an interview session first. If they will be called one by one to enter the interview room, then to calculate the chances of each candidate for student council chairman being called does not have to be based on the initial conditions.

What do you think of the two statements above?

A. Statement one is true and statement two is false.
   Statements one and two should be conditional probabilities because to calculate the probability, the initial conditions must be considered.
B. Statement one is incorrect and statement two is correc.
   Statements one and two should be unconditional probabilities because to calculate the probability, you don't have to consider the initial conditions.
C. Statement one and statement two are true.

EDUMATIKA: Jurnal Riset Pendidikan Matematika *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Statement one is a conditional probability because to calculate the probability, the initial conditions must be considered. While statement two is an unconditional probability because to calculate the probability does not have to pay attention to the initial conditions.

D. Statement one and statement two are wrong.
Statement one is an unconditional probability because to calculate the probability, you don't have to pay attention to the initial conditions. While statement two is a conditional probability because to calculate the probability it must pay attention to the initial conditions.

E. Statement one or statement two is false.
Statement one is a conditional probability because to calculate the probability, the initial condition must be considered. While statement two is an unconditional probability because to calculate the probability it does not have to pay attention to the initial conditions.

Table 6. Example Item 2 (Applying Level)

Level	6 (Grade 11 and 12)
Domain	Data and Uncertainty
Subdomain	Sample space of events and combinations
Context	Social Culture
Competency	Determining combinations to calculate the sample space of events
Cognitive Level	Applying
Problem Form	Multiple Choice

Kebaya is an Indonesian cultural heritage that must be preserved. In this modern era, kebaya is often used as a dresscode for any event, be it graduation, organization, office, or wedding. As a result, there are kebayas that are designed to be as comfortable as possible so that they can be used for daily activities. Toko D'Lote is a shop that sells various types of kebaya, but with a limited total production of 214 kebaya every day. The production details at Toko D'Lote are as follows:

Types of Kebaya	Children	Adults
Modern Kutubaru Kebaya	36	48
Kebaya with Cape (Long Shawl)	36	24
Kebaya with Puff Sleeves	-	30
Transparent Kebaya	-	18
Kebaya with Rich Embroidery	-	36

On that day, Daisy became the seventh customer in D'Lote's shop with details of the six previous customers buying products:

Customer	Clothing
1	1 Modern Kutubaru Kebaya (Children), 1 Modern Kutubaru Kebaya (Adults)
2	2 Kebaya with Puff Sleeves (Adult)
3	2 Kebaya with Cape (Long Shawl) (Children), 1 Kebaya with Rich Embroidery (Adults)
4	1 Kebaya with Cape (Long Shawl) (Children), 1 Kebaya with Cape (Long Shawl) (Adults), 2 Kebaya with Puff Sleeves (Adults)
5	1 Transparent Kebaya (Adult)
6	1 Modern Kutubaru Kebaya (Adult), 1 Transparent Kebaya (Adult)

If Daisy buys 2 kebayas, the probability that on that day she buys 2 transparent kebayas is...% (rounding to two numbers behind the comma)
A. 0,55 %
B. 0,54 %
C. 0,53 %
D. 0,52 %
E. 0,51 %

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Table 7. Example Item 3 (Reasoning Level)

Level	6 (Grade 11 and 12)
Domain	Data and Uncertainty
Subdomain	Independent events and probability calculation
Context	Personal
Competency	Analyzing two independent events based on probability calculations
Cognitive Level	Reasoning
Problem Form	Multiple Choice

SMAN 1 Pekanbaru has a school canteen. The school canteen has 12 types of food and 8 types of drinks and sells food and drinks separately, meaning that the choice of food does not affect the choice of drinks. A student wants to buy two foods and one drink. If the food and drink are chosen at random. What is the probability of a particular food and a particular drink being chosen? Based on the calculated probabilities, are the events "choosing a certain food" and "choosing a certain drink" mutually independent based on the combination calculation?

A. $\frac{1}{528}$, the events are mutually independent because the probabilities of food and drink do not affect each other.
B. $\frac{1}{528}$, the events are not mutually independent because the purchase is made without returns.
C. $\frac{1}{528}$, the events are mutually independent because the probability of food and drink affecting each other.
D. $\frac{1}{528}$, the events are mutually independent only if the amounts of food and drink are equal.
E. $\frac{1}{528}$, these events affect each other only when the amount of food or drink in the cafeteria changes.

The content validity of the nine items was tested by three experts in mathematics education using Aiken's V method. Items with Aiken's V ≥ 0.80 were considered valid, while those in the moderate range ($0.40 < V \leq 0.80$) required partial revision based on weak indicators. None of the items fell below the minimum validity threshold, but several indicators within items required refinement based on expert input. Opinions and input from experts are carefully evaluated and applied to the questions developed (Susantini et al., 2025). The expert feedback focused primarily on contextual alignment and clarity of wording. Table 8 shows an example of item revision following expert feedback.

EDUMATIKA: Jurnal Riset Pendidikan Matematika *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Table 8. Example of Item Revision Based on Expert Feedback

Cognitive Level	Problem Before Revision	Some Expert Assessment Comments	Problem After Revision
Reasoning	 SMAN 1 Pekanbaru has a school canteen. The school canteen has 12 types of food and 8 types of drinks and sells food and drinks separately, meaning that the choice of food does not affect the choice of drinks. A student wants to buy two foods and one drink. If the food and drink are chosen at random. What is the probability of a particular food and a particular drink being chosen? Based on the calculated probabilities, are the events "choosing a certain food" and "choosing a certain drink" mutually independent based on the combination calculation?	The picture does not match the context. Look at the competencies again.	SMAN 1 Pekanbaru has a school canteen that provides 12 types of food and 8 types of drinks. Food and beverages are sold separately, so the choice of food does not affect the choice of beverages. A student will buy two foods and one drink, all chosen randomly with no returns. If the student wants to get one particular food out of the two selected foods, and one particular drink, what is the probability of getting the particular food and the particular drink? Based on the results of the probability calculation, can the events "choosing a certain food" and "choosing a certain drink" be categorized as mutually independent events?

Content Validity of the MCA-Equivalent Summative Assessment

The content validity analysis using Aiken's V produced indices ranging from 0.64 to 0.73, indicating moderate validity according to the Aiken category (Susantini et al., 2025). Table 9 presents a snapshot of the results of the instrument validity for each cognitive level and indicator.

Table 9. Summary of Content Validity Results by Cognitive Level

Cognitive Level	Indicator	Expert 1	Expert 2	Expert 3	Validity Index
Knowing	1	3	4	4	0,67
	2	4	4	4	0,75
	3	3	4	4	0,67
	4	3	4	4	0,67
	5	3	4	4	0,67
	6	3	4	3	0,58
	7	4	2	3	0,5
	8	4	4	3	0,67
	9	4	4	4	0,75
	10	2	4	4	0,58
	11	4	3	4	0,67
	12	3	4	3	0,58
	13	4	4	5	0,83
	14	4	4	2	0,58
	15	4	4	3	0,67
	16	3	4	3	0,58

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

Cognitive Level	Indicator	Expert 1	Expert 2	Expert 3	Validity Index
	17	4	4	3	0,67
	1	4	4	4	0,75
	2	4	4	4	0,75
	3	4	4	3	0,67
	4	4	4	4	0,75
	5	4	4	4	0,75
	6	4	4	4	0,75
	7	4	2	4	0,58
	8	4	4	4	0,75
Applying	9	3	4	5	0,75
	10	5	4	4	0,83
	11	4	3	4	0,67
	12	3	4	5	0,75
	13	4	4	5	0,83
	14	4	4	4	0,75
	15	4	4	4	0,75
	16	3	4	4	0,67
	17	4	4	3	0,67
	1	4	3	4	0,67
	2	3	3	4	0,58
	3	3	4	3	0,58
	4	4	4	3	0,67
	5	4	3	4	0,67
	6	3	3	3	0,5
	7	4	2	3	0,5
	8	4	4	3	0,67
Reasoning	9	3	4	5	0,75
	10	2	4	3	0,5
	11	3	3	4	0,58
	12	4	4	4	0,75
	13	3	4	5	0,75
	14	4	4	4	0,75
	15	4	4	4	0,75
	16	4	4	4	0,75
	17	3	4	3	0,58

Based on the results of the content validity the average Aiken's V index for the summative assessment instrument for Combinatorics material equivalent to MCA is 0.68, so this instrument falls into the moderate validity category (Susantini et al., 2025). This means that the instrument still

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

requires refinement to fully align with the intended cognitive levels, contexts, and competencies. Therefore, following the expert feedback, each item will undergo revision, especially on indicators with low Aiken indices, to improve clarity, contextual accuracy, and cognitive demand. Revision process is needed for each item to ensure its usefulness in future assessments (An Nabil et al., 2022).

A valid assessment instrument should reflect content that is consistent with learning objectives and accurately measures students' comprehension of mathematical concepts. A high quality assessment will be able to accurately measure the learning objectives to be achieved because it involves examining each question to determine whether it aligns with the competencies being measured (Pandra & Aswarliansyah, 2023). Therefore, enhancing content validity ensures that the instrument can later function as a reliable evaluation tool for learning outcomes. This is in line with the fact that content validity is a function of how well the dimensions and elements of a concept have been described (Roebianto et al., 2023).

The Aiken's V index of items with cognitive level of applying has the highest Aiken index than items with cognitive level of knowing and reasoning. Aiken's index of items with cognitive level of applying is between 0.67 to 0.73. While items with cognitive level of knowing have Aiken index between 0.64 to 0.70. And items with cognitive level reasoning have Aiken index between 0.65 to 0.70. This suggests that assessment items requiring procedural application are easier to construct and validate than those assessing conceptual understanding or reasoning. In the process of learning mathematics, students usually apply the concept of memorization, so that students are more familiar with questions that apply formulas (Izzati et al., 2023). When faced with questions that test their understanding of the application of mathematics in everyday life, students will find it difficult to explain. These results indicate that to formulate questions with cognitive level of knowing and reasoning is more difficult than formulating questions with cognitive level of applying.

One of the problems of numeracy skills that occur at the cognitive level of reasoning is the aspect of evaluating ability, namely students do not use the right strategy and cannot provide other alternative solutions to solve problems (Viyana et al., 2025). The results of the Trends in International Mathematics and Science Study (TIMSS) 2015 and Programme for International Student Assessment (PISA) 2015 show that Indonesian students still face difficulties in understanding and applying knowledge to solve complex problems, draw conclusions, and make generalizations (reasoning) (Putra et al., 2018). Therefore, teachers need to design learning experiences that gradually build reasoning ability, integrating it into everyday contexts before testing it through high-level items.

The main challenge in creating good reasoning questions is not only in the preparation of questions, but also in the readiness of students to face questions that require analysis, evaluation, and reflection (Viyana et al., 2025). On the other hand, the mistakes made by students in working on

EDUMATIKA: *Jurnal Riset Pendidikan Matematika* *e-ISSN 2620-8911*
Volume 8, Issue 2, November 2025 *p-ISSN 2620-8903*

numeracy literacy problems that knowing is that students have difficulty in working on arithmetic operations so that they make mistakes in operating numbers (Viyana et al., 2025).

The results highlight the importance of expert validation and iterative refinement in developing valid assessment instruments. Moderate validity scores suggest that while the instrument is functional, further revision will enhance its accuracy in assessing students' numeracy competencies in Combinatorics. This iterative process reflects the essence of R&D studies, where validation and revision are crucial for producing a robust educational instrument. Once improved, the instrument can serve as a standardized tool for evaluating Combinatorics learning aligned with MCA numeracy domains. The findings also have practical implications for mathematics educators and assessment developers, suggesting that integrating numeracy-oriented frameworks such as those proposed by OECD (2021) and Goos et al. (2014) can strengthen the authenticity and contextual relevance of mathematics assessments. Nevertheless, this study has certain limitations. The number of expert validators was limited to three, and the validation process focused solely on content validity without including reliability or empirical testing. According to Rios and Wells (2014), comprehensive validation should incorporate evidence from multiple sources, including internal structure and response processes, to ensure a holistic assessment of validity.

CONCLUSION

The validity of the MCA-equivalent Combinatorics summative assessment instrument using the Aiken method, the range was between 0.64 and 0.73. This means that the instrument falls into the medium validity category. Therefore, in this study, it is necessary to revise the items with low validity in order to produce a high-quality MCA-equivalent Combinatorics summative assessment instrument. Once revised, this instrument is expected to serve as a reference for educators in developing MCA-equivalent assessment questions. Where assessments not only measure procedural skills but also test higher-order thinking skills, enabling educators to design more meaningful learning evaluations. The findings of this study imply that the Aiken validity approach provides a practical and systematic framework for evaluating item quality in the development of higher-order mathematics assessments. Future research is recommended to employ complementary validation methods such as the Gregory approach with at least three expert validators, and to extend the process by conducting reliability analysis and empirical trials involving students.

REFERENCES

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings, educational and psychological measurument. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Alhunaini, S., Osman, K., & Abdurab, N. (2020). The development and validation of assessment practices of mathematical thinking (APMT) instrument (Pembangunan dan pengesahan instrumen amalan pentaksiran pemikiran matematik (APMT). *Jurnal Pendidikan Malaysia*, *45*(02), 79–91. https://doi.org/10.17576/jpen-2020-45.02-08

An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *Paedagogia*, *25*(2), 184. https://doi.org/10.20961/paedagogia.v25i2.64566

Andrianti, D. S., & Rahayu, P. (2022). Kemampuan literasi numerasi berdasarkan kecerdasan logis matematis melalui soal AKM pada siswa SMP. *JTMT: Journal Tadris Matematika*, *3*(2), 55–63. https://doi.org/10.47435/jtmt.v3i2.1189

Aydinyer, Y., & Ubuz, B. (2025). Development and Validation of deep procedural knowledge test on angles and polygons (Deep-ProKT-AngPoly). *International Journal of Science and Mathematics Education*, *0123456789*. https://doi.org/10.1007/s10763-025-10543-0

Batanero, C., Navarro-Pelayo, V., & Godino, J. D. (1997). Effect of the implicit combinatorial model on combinatorial reasoning in secondary school pupils. *Educational Studies in Mathematics*, *32*, 181–199. https://doi.org/10.1023/A:1002954428327

Chytry, V., & Kubiatko, M. (2021). Pupils' summative assessments in mathematics as dependent on selected factors. *Eurasia Journal of Mathematics, Science and Technology Education*, *17*(8), 1–11. https://doi.org/10.29333/EJMSTE/11112

Goos, M., Geiger, V., & Dole, S. (2014). Transforming professional practice in numeracy teaching. In Y. Li, E. A. Silver, & S. Li (Eds.), *Transforming Mathematics Instruction: Multiple Approaches and Practices* (pp. 81–102). Springer. https://doi.org/10.1007/978-3-319-04993-9_6

Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Springer. https://doi.org/10.1007/978-94-007-2324-5

Hilden, R., Oscarson, A. D., Yildirim, A., & Fröjdendahl, B. (2022). Swedish and Finnish pre-service teachers' perceptions of summative assessment practices. *Languages*, *7*(1), 1–23. https://doi.org/10.3390/languages7010010

Izzati, N., Wahyuni, N., & Irawan, B. (2023). Development of an interactive PowerPoint to instill a deeper understanding of polyhedron concepts in mathematics class. *Alifmatika: Jurnal Pendidikan Dan Pembelajaran Matematika*, *5*(1), 95–109. https://doi.org/10.35316/alifmatika.2023.v5i1.95-109

Kania, N., Kusumah, Y. S., Dahlan, J. A., Nurlaelah, E., Gürbüz, F., & Bonyah, E. (2024). Constructing and providing content validity evidence through the Aiken's V index based on the experts' judgments of the instrument to measure mathematical problem-solving skills. *REID (Research and Evaluation in Education)*, *10*(1), 64–79. https://doi.org/10.21831/reid.v10i1.71032

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lestari, P., & Ali, N. N. (2023). Website-based learning: Optimizing students' numeracy skills. *Edumatika : Jurnal Riset Pendidikan Matematika*, *6*(2), 82–91. https://doi.org/10.32939/ejrpm.v6i2.2973

Lukman, H. S., & Setiani, A. (2018). Validitas bahan ajar statistika terapan berbasis ICT terintegrasi proyek. *Edumatika : Jurnal Riset Pendidikan Matematika*, *1*(2), 36. https://doi.org/10.32939/ejrpm.v1i2.226

Nurjanah, S., Istiyono, E., Widihastuti, W., Iqbal, M., & Kamal, S. (2023). The application of Aiken's V method for evaluating the content validity of instruments that measure the implementation of formative assessments. *Journal of Research and Educational Research Evaluation*, *12*(2), 125–133. http://journal.unnes.ac.id/sju/index.php/jere

OECD. (2023). *PISA 2022 Assessment and Analytical Framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

Pandra, V., & Aswarliansyah. (2023). Validitas isi instrumen tes matematika siswa kelas VII SMP semester ganjil. *Jurnal Pendidikan Matematika: Judika Education*, *13*(1), 104–116. https://doi.org/10.31539/judika.v6i2.7631

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure what's being reported? Critique

and recommendations. *Research in Nursing & Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Prafianti, R. A., Ilmayasinta, N., & Silvia, I. A. (2023). Validitas buku ajar matematika SMP sebagai penguat asesmen kompetensi minimum dengan konteks sosial budaya Jawa Timur. *JIPMat*, *8*(2), 210–219. https://doi.org/10.26877/jipmat.v8i2.16380

Pratiwi, I. R., Sari, E. M., Novitasari, N., Zikri, A., & Putra, M. G. P. (2023). Pengembangan aplikasi Smart – Calculator berbasis digital sebagai media pembelajaran materi kombinatorika [Developing Smart application – Digital-based Calculator as a reinforcement media in learning discrete mathematics]. *JOHME: Journal of Holistic Mathematics Education*, *7*(2), 180. https://doi.org/10.19166/johme.v7i2.7450

Putra, A., Syarifuddin, H., & Zulfah. (2018). Validitas lembar kerja peserta didik berbasis penemuan terbimbing dalam upaya meningkatkan pemahaman konsep dan kemampuan penalaran matematis. *Edumatika : Jurnal Riset Pendidikan Matematika*, *1*(2), 56–62. https://doi.org/10.32939/ejrpm.v1i2.302

Putri, M. R., As'ari, A. R., & Sisworo, S. (2020). Gesture mahasiswa dalam menyelesaikan masalah kombinatorika dengan jawaban benar berdasarkan tahap berpikir Mason. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *5*(8), 1190. https://doi.org/10.17977/jptpp.v5i8.13962

Putri, Y. F., Kadir, K., & Dimyati, A. (2022). Analysis of content validity on mathematical computational thinking skill test for junior high school student using Aiken method. *Hipotenusa : Journal of Mathematical Society*, *4*(2), 108–119. https://doi.org/10.18326/hipotenusa.v4i2.7465

Ramadani, A. H., Ekohariadi, Anifah, L., Nugroho, Y. S., & Safitri, R. (2024). Development of performance assessment instruments for measuring drawing skills in vocational students competence in mechanical engineering expertise. *IJORER : International Journal of Recent Educational Research*, *5*(2), 268–279. https://doi.org/10.46245/ijorer.v5i2.551

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, *26*(1), 108–116. https://doi.org/10.7334/psicothema2013.260

Roebianto, A., Savitri, S. I., Aulia, I., Suciyana, A., & Mubarokah, L. (2023). Content validity: Definition and procedure of content validation in psychological research. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, *30*(1), 5–18. https://doi.org/10.4473/TPM30.1.1

Siahaan, M. M. L., Hijriani, L., & Toni, A. (2022). Identifikasi kemampuan literasi numerasi melalui instrumen asesmen kompetensi minimum pada siswa SMA kelas XI SMAS Warta Bakti Kefamenanu [Identification of the numerical literacy ability of grade 11 students at Warta Bakti Kefamenanu high school using the minimum competency assessment instrument. *JOHME: Journal of Holistic Mathematics Education*, *6*(2), 178. https://doi.org/10.19166/johme.v6i2.5751

Susantini, E., Sari, Y. M., Asteria, P. V., & Marzuqi, M. I. (2025). Proving content validity of Android-based higher order thinking skill assessment for science and mathematics preservice teacher. *Journal of Education and Learning*, *19*(1), 551–560. https://doi.org/10.11591/edulearn.v19i1.21207

Taherdoost, H. (2016). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in research. *International Journal of Academic Research in Management*, *5*(3), 28–36. https://doi.org/10.2139/ssrn.3205040

Tandas, J. B. (2021). Development of a valid and reliable summative test in plane trigonometry. *Journal of Science and Mathematics Letters*, *9*(1), 45–59. https://doi.org/10.37134/jsml.vol9.1.5.2021

Viyana, D., Mardiyana, Riyadi, & Usodo, B. (2025). Kesalahan siswa dalam menyelesaikan soal trigonometri ditinjau dari literasi numerasi pada level konitif knowing, applying, dan reasoning. *JIPMat (Jurnal Ilmiah Pendidikan Matematika)*, *10*(1), 31–42. https://doi.org/10.26877/jipmat.v10i1.1645

Zakiyah, Z., & Kartika, H. (2024). Uji validitas konten instrumen kemampuan representasi matematis dalam menyelesaikan masalah bangun datar. *JP2M (Jurnal Pendidikan Dan Pembelajaran Matematika)*, *10*(1), 250–257. https://doi.org/10.29100/jp2m.v10i1.5485