



Validity of Items and Spurious Overlap (ESO) Effects Unidimensionally and Multidimensionally (A Study of the Effects of Spurious Overlap on the Test of Logical Thinking Ability)

I Gusti Ngurah Puger^{1*} 

¹ Education Department, Faculty of Teacher Training and Education, Universitas Panji Sakti, Singaraja, Indonesia

ARTICLE INFO

Article history:

Received November 03, 2023

Accepted February 04, 2024

Available online February 25, 2024

Kata Kunci :

Validitas Butir, Spurious Overlap, Unidimensi, Multidimensi, Kemampuan Berpikir Mantik

Keywords:

Item Validity, Spurious Overlap, Unidimensionality, Multidimensionality, Logical Thinking Ability



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright ©2024 by Author. Published by Universitas Pendidikan Ganesha

ABSTRAK

Pada tingkat SMP masih sedikit sekali tes kemampuan berpikir mantik yang sudah distandarisasi untuk digunakan mengukur kemampuan berpikir mantik siswa. Tujuan dari penelitian ini untuk menganalisis validitas butir dan efek spurious overlap (ESO) secara unidimensi dan multidimensi (suatu kajian efek spurious overlap pada tes kemampuan berpikir mantik. Penelitian ini termasuk jenis penelitian kalibrasi instrumen. Populasi penelitian ini adalah seluruh siswa kelas VIII SMP dengan sampelnya berupa jawaban 100 siswa yang dipilih secara purposive sampling. Metode pengumpulan data dengan tes. Hasil analisis data diperoleh temuan pada pengujian validitas butir tes kemampuan berpikir mantik secara unidimensi sebelum dikoreksi dengan efek spurious overlap, terdapat 49 butir yang layak digunakan sebagai penyusun tes; dan pengujian validitas butir tes kemampuan berpikir mantik secara multidimensi sebelum dikoreksi dengan efek spurious overlap, terdapat 49 butir yang layak digunakan sebagai penyusun tes, pada pengujian validitas butir secara unidimensi setelah dikoreksi dengan efek spurious overlap, terdapat 47 butir yang layak digunakan sebagai penyusun tes; dan pengujian validitas butir tes kemampuan berpikir mantik secara multidimensi sebelum dikoreksi dengan efek spurious overlap, terdapat 45 butir yang layak digunakan sebagai penyusun tes, dan koefisien ESO secara unidimensi lebih kecil bila dibandingkan dengan koefisien ESO secara multidimensi dari butir yang menyusun tes kemampuan berpikir mantik.

ABSTRACT

At the junior high school level, there are still very few standardized tests of logical thinking abilities that can be used to measure students' logical thinking abilities. This research aims to analyze the validity of items and the effect of spurious overlap (ESO) unidimensionally and multidimensionally (a study of the effect of spurious overlap on tests of logical thinking abilities). This research is a type of instrument calibration research. The population of this research is all students of class VIII at junior high school, with the sample in the form of answers from 100 students' selected using purposive sampling. Data collection method using tests. The results of data analysis obtained findings from testing the validity of unidimensional logical thinking ability test items before being corrected for the spurious overlap effect; 49 items were suitable for use as part of the test, and testing the validity of the multidimensional logical thinking ability test items before being corrected for the spurious overlap effect, 49 items are suitable for use as test constructors, in the unidimensional test of item validity after being corrected for the spurious overlap effect, 47 items are suitable for use as test constructors; and testing the validity of the multidimensional logical thinking ability test items before being corrected for the spurious overlap effect, 45 items are suitable to be used as test components, and the unidimensional ESO coefficient is smaller when compared to the multidimensional ESO coefficient of the items that make up the logical thinking ability test.

1. INTRODUCTION

The ability to think logically is a thinking process that combines deductive thinking with inductive thinking. The truth of the premises in deductive thinking must be tested theoretically, so deductive thinking is often known as rational thinking (Karjo et al., 2019; Xu et al., 2022). Deductive thinking is a process of drawing conclusions starting from the major premise leading to the minor premise (Leonard et al., 2023). This deductive thinking can be used as a means to carry out the process of formalizing concepts (Bhaw et al., 2023; Voskoglou & Broumi, 2022). Meanwhile, inductive thinking is a decision-making process that starts from specific facts and leads to general statements (Talman et al., 2021). Inductive thinking can be used as a means to concretize concepts. Deductive logic and inductive logic are combined in a thinking

*Corresponding author.

E-mail addresses: flugone@yahoo.com (I Gusti Ngurah Puger)

concept, then the ability to think logically is obtained (Abate et al., 2020). Logical thinking can be applied in terms of understanding concepts in the field of science studies. Deductive thinking is used as a means to understand concepts abstractly and inductive thinking is used as a means to understand concepts concretely. Students who have high logical thinking abilities tend to have high memory, and also have higher activity in the learning process in class when compared to students who have low logical thinking abilities (Khasanah et al., 2020). So, the level of students' logical thinking abilities greatly determines their biology learning outcomes.

Currently, in science learning in Junior High Schools, it has been a tradition that has been in effect for generations, that teaching material is delivered conventionally. What this means conventionally is that teachers in the fields of biology and physics use the lecture method in delivering their teaching material to students in class. Books that are deemed suitable for use as teaching materials are delivered by the teacher to be purchased at the nearest bookstore. In fact, it is not uncommon for science teachers to buy the necessary textbooks and then distribute them to students (Purnama Dewi, 2018; Septriningsih et al., 2010). The follow-up is that students pay the price of the books distributed to the teacher concerned. Students are relatively forced to read books that are used as reference materials, with the hope that apart from being explained at school, students can study the material in their respective homes. It is rare for students to be invited into the field to observe nature directly, in connection with materials that need to be communicated to students. The environment is a very good and complete natural laboratory, but not many people realize and use it (Darmayasa et al., 2018; Kurnia et al., 2021; Nuraeni & Habibi, 2021). Apart from that, at the junior high school level there are still very few tests of logical thinking abilities that have been standardized to be used to measure students' logical thinking abilities. The logical thinking ability test was tested for content validity, inter-rater response reliability, item validity testing, and test reliability calculations. In terms of testing the validity of the items, most tests of logical thinking abilities only report testing unidimensionally. In fact, the test of logical thinking ability, if seen from the grid made by science teachers, contains more than one dimension. This means that the test of logical thinking ability must report testing the validity of items in a multidimensional manner. Testing the validity of the unidimensional logical thinking ability test items uses the product moment correlation formula without being corrected for the effect of spurious overlap on the total score. This incident causes a correlation coefficient (r_{ix}) that is higher (overestimate) than the actual correlation.

The solution to measure students' logical thinking abilities is to use standard logical thinking ability tests. The logical thinking ability test items that are tested for unidimensional validity, some of which are categorized as valid or dropped, essentially still contain spurious overlap effects. Testing the validity of unidimensional logical thinking ability test items using the product moment correlation formula should be corrected for the spurious overlap effect (Puger & Dewi, 2023). Test scores are generally just the sum of the scores of all the items. When there are not enough items in the test being analyzed, the item-total correlation coefficient (r) obtained will tend to be high and is an overestimation of the actual correlation coefficient (Khasanah et al., 2020; Lestari & Harjono, 2021). This overestimation occurs because of the large contribution of each item in determining the test score (Baskara & Yudiana, 2021). The score for each item becomes part or portion of the test score. The quantity of this portion will be greater if the number of items in the test is smaller (Larasati & Widyastuti, 2020).

Previous research findings stated that by testing the validity of the unidimensional and multidimensional logical thinking ability test, which was carried out before and after being corrected for the spurious overlap effect, we would be able to calculate the magnitude of the spurious overlap (ESO) effect value on the test score (total score) for each item (Warju et al., 2020). By knowing the magnitude of the spurious overlap (ESO) effect value, the concept of the spurious overlap effect on the total score can be concretized, or can be understood concretely (Puger & Dewi, 2023). It is hoped that the results of this research will be able to add insight into the field of preparing research instruments, especially instruments in the form of tests. Another important outcome of this research is the production of a test of logical thinking abilities which has been assessed for the validity of its items both unidimensionally and multidimensionally, both before and after correction for the spurious overlap effect. Apart from that, this test of logical thinking ability will be equipped with a description of the magnitude of the contribution of the spurious overlap effect to the total score both unidimensionally and multidimensionally, after correction for the spurious overlap effect. The aim of this research is to analyze the validity of items and the effect of spurious overlap (ESO) unidimensionally and multidimensionally (a study of the effect of spurious overlap on tests of logical thinking abilities).

2. METHODS

This research is a type of instrument calibration research. Instrument calibration research is a type of research by conducting internal and external analysis studies of an instrument, so that a standardized instrument is formed (Paramartha et al., 2022). This research carried out the process of testing the validity of the items and ESO coefficients of the research instrument in the form of a 'logical thinking ability test.' External analysis was carried out through unidimensional and multidimensional item validity tests, both before and after correction for the spurious overlap effect. This research will also reveal the magnitude of the unidimensional and multidimensional spurious overlap (ESO) effect coefficient, after correction for the spurious overlap effect. From these six units of analysis, a decision can be made regarding the suitability of the logical thinking ability test to be used as a research instrument.

The population of this research is the answers to the logical thinking ability test developed by the researcher. Meanwhile, the excerpt is data on the answers of 100 class VIII students of SMP Negeri 1 Seririt to the logical thinking ability test in the form of answers to each item on the logical thinking ability test. Determining the subjects as many as 100 class VIII students in this study used a purposive sampling technique. Purposive sampling technique is a way of taking samples in accordance with previously determined research objectives (Agung, 2018). Sample members are selected from subgroups in the population that suit the research objectives. Data collection method using tests. The instrument needed to obtain data in the form of test answers is a logical thinking ability test developed by researchers. The test consists of 54 items and must be answered by 100 students within 60 minutes. Before this test was prepared, an operational definition of logical thinking ability and a technical thinking ability test specification table were first created, which contained variable names, dimensions (aspects), indicators and item numbers. The instrument grid is a test of logical thinking abilities as listed in Table 1.

Table 1. Logic Thinking Ability Test Grid

Topic	Subtopic	Indicator	Amount Item	Number Item	
The Ability to Think Logically	1. Deductive thinking.	Students can:			
		1. Major premise	1) Explain the definition of the major premise correctly (C1)	2	
			2) Identify the major premise in a deductive thinking building (C4)	2	10, 40
			3) Mention the role of the major premise in building deductive thinking (C1)	2	26, 54
	2. Minor premise		4) Explain the definition of the minor premise correctly (C1)	2	4, 20
			5) Identify the minor premise in a deductive thinking building (C4)	2	12, 29
			6) Mention the role of the minor premise in building deductive thinking (C1)	2	39, 46
	3. Conclusion		7) Explain the definition of conclusion correctly (C1)	2	1, 30
			8) Identify conclusions in a deductive thinking building (C4)	2	13, 23
			9) Compare the conclusion with the major premise based on the meaning contained (C2)	2	34, 52
	4. Middle term (term medius)		10) Explain the definition of the middle term correctly (C1)	2	8, 51
			11) Identify the middle term in the major premise (C4)	2	16, 50
			12) Identify the middle term in the minor premise (C4)	2	21, 43
5. Means of concept abstraction		13) Explain the definition of abstract (formal) concepts correctly (C1)	2	9, 37	

Topic	Subtopic	Indicator	Amount Item	Number Item
		14) Identify 1 example of an abstract concept (C4)	2	27, 49
	II. Inductive thinking			
	a. Special statements	15) Explain the definition of a special statement correctly (C1)	2	2, 53
		16) Mention 1 example of a special statement (C1)	2	14, 28
		17) Mention the role of special statements in inductive thinking (C1)	2	31, 48
	b. The knot	18) Explain the definition of conclusions in inductive thinking (C1)	2	19, 41
		19) Identify conclusions in inductive thinking (C4).	2	25, 33
	c. Principle of analogy	20) Explain the definition of the principle of analogy (C1)	2	7, 44
		21) Identify the principle of analogy in building an induction analogy (C4)	2	17, 38
		22) Mention the role of the principle of analogy in understanding concepts (C1)	2	32, 45
	d. The conclusion of the induction analogy	23) Define the conclusion of the induction analogy correctly (C1)	2	5, 24
		24) Identify the conclusion of the induction analogy (C4)	2	3, 18
	e. Means of concretization	25) Explain the definition of concrete concepts correctly (C1)	2	11, 22
		26) Provide reasons for inductive thinking as a means of concretizing concepts (C1).	2	35, 47
		27) Identify 1 example of a concrete concept (C4)	2	36, 42
Amount			54	

Based on the test grid listed in Table 1, a logical thinking ability test can finally be compiled, consisting of 54 items. For more details regarding the logical thinking ability test, please review Appendix 1 page 94. The logical thinking ability test used to collect data in this research has been tested for content validity and the inter-rater response reliability has been calculated. By using Gregory's formula, a content validity coefficient (VI) of 0.98 was obtained. Meanwhile, the reliability of the inter-rater response was calculated using the Anava Hoyt formula, which then obtained an $r_{xx'}$ coefficient of 0.81. Because the VI coefficient (= 0.98) is greater than 0.90, meaning that it is assessed from its content validity, the logical thinking ability test can be used further. Likewise, if the coefficient $r_{xx'}$ (= 0.81) is greater than 0.70, it means that judging from the reliability of the inter-rater responses, the developed logical thinking ability test can be used further.

In essence, the main objective of this research is to determine the unidimensional item-total correlation coefficient (r_{ix}), the unidimensional item-total correlation coefficient corrected for the spurious overlap effect ($r_{i(xi)}$), the spurious overlap effect coefficient (ESO).) unidimensionally, item-total correlation coefficient (r_{ix}) multidimensionally, item-total correlation coefficient corrected for spurious overlap effect ($r_{i(xi)}$) multidimensionally, and coefficient spurious overlap effect (ESO) multidimensionally from the logical thinking ability test. To test the validity of the logical thinking ability test items before being corrected for the spurious overlap effect, both unidimensionally and multidimensionally, the product moment correlation formula was used. The test criteria used are comparing the r_{ix} price with the critical table price r , with the condition that r_{ix} is said to be valid if $r_{ix} > r$ -table at a significance level of 5%. With a number of testees (N) of 100 and a significance level of 5%, the r -table value is 0.195. Thus, it can be said

that if the rix value is > 0.195 , then the rix value is in the valid category. On the other hand, if the rix value is < 0.195 , then the rpb value is included in the drop category. To test the validity of the logical thinking ability test items after being corrected for the spurious overlap effect, both unidimensionally and multidimensionally, the corrected product moment correlation formula was used.

The criterion used is to compare the price of $r_{i(xi)}$ -calculated with the critical table value r , provided that $r_{i(xi)}$ -calculated is said to be valid if $r_{i(xi)}$ -calculated $> r$ -table at the 5% significance level. With a number of testees (n) of 100 and a significance level of 5%, the r -table value is 0.195. Thus, it can be said that if the $r_{i(xi)}$ -calculation value is > 0.195 , then the $r_{i(xi)}$ -calculation value is in the valid category. On the other hand, if the $r_{i(xi)}$ -calculation value is < 0.195 , then the $r_{i(xi)}$ -calculation value is included in the drop category. To determine the contribution of the spurious overlap effect to the total score, both unidimensionally and multidimensionally (ESO coefficient), the formula is used: $ESO = rix - r_{i(xi)}$. Where, ESO is the spurious overlap effect, rix is the product moment correlation coefficient, and $r_{i(xi)}$ is the correlation coefficient corrected for the spurious overlap effect. The aim of determining the ESO coefficient is to determine the contribution of the spurious overlap effect to the total score (test score) which causes the rix coefficient to be overestimated when compared with the $r_{i(xi)}$ coefficient, both unidimensionally and multidimensionally.

3. RESULT AND DISCUSSION

Results

Data regarding the answers of 100 class VIII students of SMP Negeri 1 Seririt (to test the validity of items before and after being corrected for the unidimensional spurious overlap effect, to calculate the unidimensional ESO coefficient, to test the validity of the items before and after being corrected for the multidimensional spurious overlap effect, and to calculate the ESO coefficient multidimensionally). Data regarding student answers, which then produces unidimensional and multidimensional rix coefficients, and unidimensional and multidimensional $r_{i(xi)}$ coefficients are used as a basis for determining which items are categorized as valid or dropped unidimensionally and multidimensionally. This student answer data is also used to produce unidimensional and multidimensional ESO coefficients. Next, the unidimensional and multidimensional ESO coefficients are used as a basis for determining the contribution of the spurious overlap effect to the unidimensional and multidimensional total score.

The first finding, testing the validity of the logical thinking ability test items unidimensionally before correcting for spurious overlap effects. If you look at the results of content validity testing and calculating inter-rater response reliability, the analysis continues with testing the validity of the unidimensional logical thinking ability test items. Testing the validity of the logical thinking ability test items unidimensionally before correcting for spurious overlap effects using the product moment correlation formula. The rix-count value obtained for each item of the logical thinking ability test is compared with the r -table value. The r -table value for the sample size (n) is 100 subjects and the 5% significance level is 0.195. If the rix-calculation value is > 0.195 , then the questionnaire item is in the valid category, conversely if the rix-calculation value is < 0.195 , then the questionnaire item is in the dropped category. Results of testing the validity of unidimensional logical thinking ability test items before correction for the spurious overlap effect using the product moment correlation formula. The validity of unidimensional logical thinking ability test items before being corrected for the spurious overlap effect can be divided into two groups, namely: (1) test items that fall into the valid category, namely test item numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, and 54, and (2) test items that are included in the drop category, namely test item numbers: 11, 14, 15, 25, and 53. Based on the results of testing the validity of the unidimensional logical thinking ability test items before being corrected for the spurious overlap effect, it can be said that as many as 49 statement items are suitable for use as part of the logical thinking ability test, and as many as 5 items statement declared invalid.

The second finding, testing the validity of the logical thinking ability test items unidimensionally after correcting for spurious overlap effects. As a result of content validity testing and calculating inter-rater response reliability, the analysis continued with testing the validity of the unidimensional logical thinking ability test items. Testing the validity of the unidimensional logical thinking ability test items after being corrected for spurious overlap effects using the corrected product moment correlation formula. The $r_{i(xi)}$ -count value obtained for each item of the logical thinking ability test is compared with the r -table value. The r -table value for the sample size (n) is 100 subjects and the 5% significance level is 0.195. If the $r_{i(xi)}$ -count value is > 0.195 , then the test item is in the valid category, conversely if the $r_{i(xi)}$ -count value is < 0.195 , then the test item is in the dropped category. Results of testing the validity of unidimensional logical thinking ability test items after correction for the spurious overlap effect using the corrected product

moment correlation formula. The results of the analysis show that the validity of the unidimensional logical thinking ability test items after being corrected for the spurious overlap effect can be divided into two groups, namely test items that are in the valid category, namely test item numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, and 54. Test items that are included in the drop category, namely test items number: 11, 14, 15, 16, 17, 25, and 53. Based on the results of testing the validity of the unidimensional logical thinking ability test items after being corrected for the spurious overlap effect, it can be said that as many as 47 statement items were suitable for use as part of the logical thinking ability test, and as many as 7 statement items were declared invalid.

Third Finding, calculating the unidimensional spurious overlap (ESO) effect coefficient. Calculation of the unidimensional spurious overlap (ESO) effect coefficient can be done using the formula: $ESO = r_{ix} - r_i(x_i)$. In Table 2, the r_{ix} coefficient has been obtained for each unidimensional logical thinking ability test item before correction for the spurious overlap effect, while in Table 3 the $r_i(x_i)$ coefficient for each unidimensional logical thinking ability test item has been obtained after correction for the spurious effect overlap. After calculating the difference between r_{ix} and $r_i(x_i)$, the ESO coefficient is obtained. The results of the analysis can be seen that the unidimensional coefficient r_{ix} before being corrected for the spurious overlap effect is greater than the unidimensional coefficient $r_i(x_i)$ after being corrected for the spurious overlap effect. This means that the unidimensional r_{ix} before being corrected for the spurious overlap effect still contains spurious overlap effect contamination on each item as large as the ESO coefficient. So the unidimensional r_{ix} coefficient before being corrected for spurious overlap effects is always an overestimate of the true correlation value. The actual value of the r_{ix} coefficient in unidimensional item validity testing is the unidimensional coefficient $r_i(x_i)$ which is corrected for the spurious overlap effect. This is what causes the coefficient value $r_i(x_i)$ to always be underestimated when compared with the unidimensional value of the coefficient r_{ix} . The magnitude of the contribution of the spurious overlap effect to the unidimensional r_{ix} value of the logical thinking ability test items number: 1 is 0.036, 2 is 0.038, 3 is 0.048, 4 is 0.046, 5 is 0.049, 6 is 0.049, 7 is 0.050, 8 is 0.040, 9 of 0.047, 10 of 0.047, 11 of 0.054, 12 of 0.047, 13 of 0.048, 14 of 0.052, 15 of 0.053, 16 of 0.052, 17 of 0.052, 18 of 0.050, 19 of 0.048, 20 of 0.046, 21 is 0.049, 22 amounted to 0.046, 23 amounted to 0.048, 24 amounted to 0.048, 25 amounted to 0.053, 26 amounted to 0.050, 27 amounted to 0.040, 28 amounted to 0.049, 29 amounted to 0.047, 30 amounted to 0.046, 31 amounted to 0.049, 32 amounted to 0.047, 33 of 0.048, 34 0.049, 35 0.041, 36 0.051, 37 0.046, 38 0.038, 39 0.046, 40 0.046, 41 0.044, 42 0.046, 43 0.044, 44 0.046, 45 0.042, 46 is 0.044, 47 is 0.048, 48 is 0.049, 49 is 0.047, 50 is 0.049, 51 is 0.040, 52 is 0.045, 53 is 0.053, and 54 is 0.045.

The fourth finding, testing the validity of the multidimensional logical thinking ability test items before correcting for spurious overlap effects. In the deductive thinking dimension, the results of testing content validity and calculating the reliability of inter-rater responses, the analysis continued with testing the validity of the logical thinking ability test items in the deductive thinking dimension. Testing the validity of the logical thinking ability test items in the deductive thinking dimension before being corrected for spurious overlap effects using the product moment correlation formula. The r_{ix} -calculation value obtained for each item of the logical thinking ability test in the deductive thinking dimension is compared with the r -table value. The r -table value for the sample size (n) is 100 subjects and the 5% significance level is 0.195. If the r_{ix} -calculation value is > 0.195 , then the test item is in the valid category, conversely if the r_{ix} -calculation value is < 0.195 , then the test item is in the dropped category. Results of testing the validity of the logical thinking ability test items in the deductive thinking dimension before correcting the spurious overlap effect using the product moment correlation formula. The results show that the validity of test items for the logical thinking ability of the deductive thinking dimension before being corrected for the spurious overlap effect can be divided into two groups, namely test items that are in the valid category, namely test item numbers: 1, 4, 6, 8, 9, 10, 12, 13, 16, 20, 21, 23, 26, 27, 29, 30, 34, 37, 39, 40, 43, 46, 49, 50, 51, 52, and 54. Test items that fall into the drop category, namely test item number: 15. Based on the results of testing the validity of test items for the logical thinking ability of the deductive thinking dimension before being corrected for the effects *terrible overlap* It can be said that as many as 27 statements are suitable for use as part of a logical thinking ability test, and 1 statement is declared invalid.

In the inductive thinking dimension, the results of testing the validity of the logical thinking ability test items in the inductive thinking dimension before being corrected for the spurious overlap effect using the product moment correlation formula, show that the validity of the logical thinking ability test items in the inductive thinking dimension before being corrected for the spurious overlap effect can be sorted out. into two groups, namely test items which are included in the valid category, namely test item numbers: 2, 3, 5, 7, 17, 18, 19, 22, 24, 28, 31, 32, 33, 35, 36, 38, 41, 42, 44, 45, 47, and 48. The test items included in the drop category are test items number: 11, 14, 25, and 53. Based on the results of testing the validity of the logical thinking ability test items in the inductive thinking dimension before being corrected for spurious

effects Overlapping it can be said that as many as 22 statement items are suitable for use as part of a logical thinking ability test, and as many as 4 statement items are declared invalid.

The fifth finding, Testing the Validity of Test Items for Multidimensional Thinking Ability After Correcting for the Spurious Overlap Effect. In the deductive thinking dimension, if you look at the results of content validity testing and calculating the reliability of inter-rater responses, the analysis continues with testing the validity of the logical thinking ability test items in the deductive thinking dimension. Testing the validity of the logical thinking ability test items in the deductive thinking dimension after being corrected for the spurious overlap effect using the corrected product moment correlation formula. The $r_i(x_i)$ -count value obtained for each item of the logical thinking ability test in the deductive thinking dimension is compared with the r-table value. The r-table value for the sample size (n) is 100 subjects and the 5% significance level is 0.195. If the r_{ix} -calculated value is > 0.195 , then the test item is in the valid category, conversely if the $r_i(x_i)$ -calculated value is < 0.195 , then the test item is in the dropped category. The results of testing the validity of the logical thinking ability test items in the deductive thinking dimension after being corrected for the spurious overlap effect using the corrected product moment correlation formula, show that the validity of the logical thinking ability test items in the deductive thinking dimension after being corrected for the spurious overlap effect can be divided into two groups. , namely test items that are included in the valid category, namely test item numbers: 1, 4, 6, 8, 9, 10, 12, 13, 20, 21, 23, 26,27, 29, 30, 34, 37, 39, 40, 43, 46, 49, 50, 51, 52, and 54. Test items included in the drop category, namely test items number: 15, and 16. Based on the results of testing the validity of test items for the logical thinking ability of the deductive thinking dimension after being corrected for the effects surprising overlap it can be said that as many as 26 statement items are suitable for use as part of a logical thinking ability test, and as many as 2 statement items are declared invalid.

The results of testing the validity of test items for the logical thinking ability of the inductive thinking dimension after being corrected for the spurious overlap effect using the corrected product moment correlation formula, it was found that the validity of the test items for the logical thinking ability of the inductive thinking dimension after being corrected for the spurious overlap effect could be divided into two groups, namely test items that are included in the valid category, namely test item numbers: 2, 3, 5, 7, 18, 19, 22, 24, 28, 31, 32, 33, 35, 38, 41, 42, 44, 45, and 47 The test items included in the drop category are test item numbers: 11, 14, 17, 25, 36, 48, and 53. Based on the results of testing the validity of test items for the logical thinking ability of the inductive thinking dimension after being corrected for the spurious overlap effect, it can be said that as many as 19 statement items were suitable for use as part of a logical thinking ability test, and 7 statement items were declared invalid.

The sixth finding, calculating the multidimensional spurious overlap (ESO) effect coefficient. Calculation of the spurious overlap (ESO) effect coefficient on the deductive thinking dimension of the logical thinking ability test can be done using the formula: $ESO = r_{ix} - r_i(x_i)$. In Table 5, the coefficient r_{ix} has been obtained for each item of the logical thinking ability test of the deductive thinking dimension before correction for the spurious overlap effect, while in Table 7 the coefficient $r_i(x_i)$ has been obtained for each item of the test item of the logical thinking ability of the deductive thinking dimension after correction for the spurious overlap effect. After calculating the difference between r_{ix} and $r_i(x_i)$, the ESO coefficient is obtained. The r_{ix} coefficient on the deductive thinking dimension before being corrected for the spurious overlap effect is greater than the $r_i(x_i)$ coefficient on the deductive thinking dimension after being corrected for the spurious overlap effect. This means that the r_{ix} in the deductive thinking dimension before being corrected for the spurious overlap effect still contains contamination from the spurious overlap effect on each item as large as the ESO coefficient. So the r_{ix} coefficient on the deductive thinking dimension before being corrected for the spurious overlap effect is always an overestimate of the true correlation value. The actual value of the r_{ix} coefficient in testing the validity of items in the deductive thinking dimension is the $r_i(x_i)$ coefficient in the deductive thinking dimension which is corrected for the spurious overlap effect. This is what causes the value of the coefficient $r_i(x_i)$ to always be underestimated when compared with the value of the coefficient r_{ix} . The magnitude of the contribution of the spurious overlap effect to the r_{ix} valuedimensions of deductive thinking, logical thinking ability test item number: 1 is 0.065, 4 is 0.082, 6 is 0.085, 8 is 0.069, 9 is 0.081, 10 is 0.081, 12 is 0.083, 13 is 0.084, 15 is 0.091, 16 is 0.091, 20 amounted to 0.082, 21 amounted to 0.085, 23 amounted to 0.081, 26 amounted to 0.086, 27 amounted to 0.073, 29 amounted to 0.078, 30 amounted to 0.081, 34 amounted to 0.085, 37 amounted to 0.084, 39 amounted to 0.077, 40 amounted to 0.082, 43 of 0.071, 46 of 0.080, 49 is 0.083, 50 is 0.086, 51 is 0.072, 52 is 0.078, and 54 is 0.078.

Calculation of the spurious overlap (ESO) effect coefficient on the inductive thinking dimension can be done using the formula: $ESO = r_{ix} - r_i(x_i)$. In Table 6, the r_{ix} coefficient has been obtained for each item of the test item of the logical thinking ability of the inductive thinking dimension before correction for the spurious overlap effect, while in Table 8 the coefficient $r_i(x_i)$ has been obtained for each item of the test

item of the logical thinking ability of the inductive thinking dimension after the correction of the spurious overlap effect. After calculating the difference between r_{ix} and $r_i(x_i)$ is the coefficient of ESO.

The results show that the coefficient r_{ix} on the inductive thinking dimension before being corrected for the spurious overlap effect is greater than the coefficient $r_i(x_i)$ on the inductive thinking dimension after correcting for the spurious overlap effect. This means that the r_{ix} in the inductive thinking dimension before being corrected for the spurious overlap effect still contains contamination from the spurious overlap effect on each item as large as the ESO coefficient. So the r_{ix} coefficient on the inductive thinking dimension before being corrected for the spurious overlap effect is always an overestimate of the true value. The actual value of the r_{ix} coefficient in testing the validity of items in the inductive thinking dimension is the coefficient ($r_i(x_i)$) on the inductive thinking dimension which is corrected for the spurious overlap effect. This is what causes the value of the coefficient $r_i(x_i)$ to always be underestimated when compared with the value of the coefficient r_{ix} . The magnitude of the contribution of the spurious overlap effect to the r_{ix} value of the inductive thinking dimension of the logical thinking ability test items number: 2 is 0.088, 3 is 0.104, 5 is 0.106, 7 is 0.105, 11 is 0.114, 14 is 0.110, 17 is 0.110, 18 is 0.108, 19 amounted to 0.103, 22 amounted to 0.098, 24 amounted to 0.099, 25 amounted to 0.109, 28 amounted to 0.106, 31 amounted to 0.104, 32 amounted to 0.100, 33 amounted to 0.098, 35 amounted to 0.089, 36 amounted to 0.112, 38 amounted to 0.082, 41 of 0.095, 42 of 0.102, 44 at 0.099, 45 at 0.094, 47 at 0.102, 48 at 0.109, and 53 at 0.113.

If we compare the unidimensional ESO coefficient value and the multidimensional ESO coefficient value, it will be clear that the comparison of the ESO coefficient value is that testing the validity of the logical thinking ability test items involves many items (unidimensionally) and which involves fewer items (multidimensionally). As for the comparison of the unidimensional and multidimensional ESO coefficient values, the unidimensional ESO coefficient value for each item on the logical thinking ability test is smaller when compared to the multidimensional ESO coefficient value. This means that the r_{ix} coefficient in unidimensional item validity testing before correction for the spurious overlap effect contains the contamination of the spurious overlap effect on the total score (test score) is smaller when compared to the r_{ix} coefficient value in multidimensional item validity testing. In other words, the contribution of the spurious overlap effect to the validity of the logical thinking ability test items before being corrected for the multidimensional spurious overlap effect is greater than unidimensional.

Discussion

The first finding in this research was that in testing the validity of the unidimensional logical thinking ability test items before being corrected for the spurious overlap effect, there were 49 items in the valid category and 5 items in the drop category. Meanwhile, in testing the validity of the unidimensional logical thinking ability test items after correction for the spurious overlap effect, there were 47 items in the valid category and 7 items in the drop category. The low number of logical thinking ability test items that fall in unidimensional and multidimensional tests carried out before and after being corrected for the spurious overlap effect is caused by the items that make up the logical thinking ability test having been derived from theory and indicators whose structure is clear when examined from theoretical studies. It should be noted that the theory and indicators used as a basis for deriving test items for logical thinking abilities were adopted. From the dimensions and indicators of the logical thinking ability test proposed by Kusuma, it is then reduced to a logical thinking ability test grid. The logical thinking ability test grid was created as a guide for deriving all the items that make up the logical thinking ability test. When traced regressively, all logical thinking ability test items have clear indicators, dimensions and variables. For example, items number 4 and 25 of the logical thinking ability test can be traced regressively to their scientific structure. Point number 4 comes from the indicator which states that it explains the definition of the minor premise correctly, from the sub dimensions of the minor premise, the dimension of deductive thinking, and the variable of logical thinking ability. Meanwhile, item number 25 comes from an indicator which states that it identifies conclusions in the building of inductive thinking, sub-dimensions of conclusions, dimensions of inductive thinking, and variables of logical thinking ability.

The second finding in this research was that in testing the validity of the multidimensional logical thinking ability test items before being corrected for the spurious overlap effect, there were 49 items in the valid category and 5 items in the drop category. Meanwhile, in testing the validity of the multidimensional logical thinking ability test items, after correction for the spurious overlap effect, there were 45 items in the valid category and 9 items in the drop category. Between the VI coefficient and the r_{xx} coefficient, the inter-rater response is directly related to the resulting r_{ix} and $r_i(x_i)$ coefficients, both unidimensionally and multidimensionally. The VI coefficient and r_{xx} ' coefficient of inter-rater responses can be used as predictors of the number of items that experience drops in unidimensional and multidimensional item validity testing, which is carried out before and after being corrected for the spurious overlap effect. The higher the value of the coefficient VI and r_{xx} ' inter-rater response results in the value of the coefficient r_{ix} and $r_i(x_i)$

unidimensionally and multidimensionally also increasing (Baskara & Yudiana, 2021; Puger & Dewi, 2023). Especially in testing the validity of test items for unidimensional and multidimensional logical thinking abilities before being corrected for the spurious overlap effect, the rix coefficient obtained is definitely an overestimate. This is because the item scores for which item validity is sought are actually included in compiling the total score (test score). This means that there is an auto correlation between the item score and itself in the total score. This is what causes the rix coefficient obtained to always be an over estimate of the actual one ($r_i(x_i)$).

The third finding in this research is that the unidimensional ESO coefficient value for each item that makes up the logical thinking ability test is smaller than the multidimensional ESO coefficient value. It seems that the ESO coefficient cannot be separated from the discussion of the item scores that make up the score *total* (test score). Testing the validity of items unidimensionally before correcting the effect of item score involvement on the total score will result in a rix coefficient value that is higher (over estimated) than the actual one. What this means is the coefficient value $r_i(x_i)$ after removing the item scores whose validity of the items is sought in the total score. The phenomenon of the contribution of item scores to the total score in determining the rix value is known as the contribution of the spurious overlap effect. Because the test score is generally just the sum of the scores of all the items, when there are not enough items in the test analyzed, the item-total correlation coefficient obtained will tend to be high and is an overestimate of the actual correlation coefficient (Erfan et al., 2020; Muluki, 2020). This overestimation occurs because of the large contribution of each item in determining the size of the test score (Janna & Herianto, 2021; Pelikan et al., 2022). The score for each item becomes part or portion of the test score. The quantity of this portion will be greater if the number of items in the test is smaller (Bashoori & Supahar, 2018; Wulandari & Radia, 2021).

Correlation coefficient *a items* with the actual test score being the correlation coefficient between the item score and the test score containing the item score itself. In other words, the item-total correlation is the correlation of the score with a part of itself which of course causes the correlation coefficient to tend to be higher than if the correlation was calculated between the item score and the test score that does not contain the item in question (Putra et al., 2018; Yusup, 2018). This condition is called spurious overlap, which produces inaccurate correlation coefficients. The fewer items there are on the scale, the greater the overlap that occurs. On the other hand, the greater the number of items in the scale, the smaller and insignificant the effects caused by spurious overlap. As a rough guide, if the number of items in the scale is more than 30 then generally the spurious overlap effect is not that large and therefore can be ignored. Meanwhile, if the number of items in the scale is less than 30 then the influence is substantial, so it needs to be considered. For this reason, so that we obtain more accurate information regarding the correlation between items and scales, we need a correction formula for the spurious overlap effect. To standardize an instrument using more than 40 items, so that the coefficient obtained is not contaminated (noising effect) by spurious overlap, it is necessary to calculate the product moment correlation using the product moment correlation formula which is corrected for the spurious overlap effect (Rezeki & Ishafit, 2017; Yugakisha & Jayanta, 2021). The greater the number of items involved in an instrument, the smaller the spurious overlap effect contribution coefficient will be. The contribution coefficient for the spurious overlap effect (ESO) is obtained by finding the difference between the rix coefficient which still contains the spurious overlap effect and the rix coefficient which has been corrected for the spurious overlap effect ($r_i(x_i)$). The implications of this research can be used as a guide for test constructors in testing the validity of test items, and the results of this research can be used as a way of testing the validity of test items to eliminate the noise effect of the spurious overlap effect on the total score.

4. CONCLUSION

The ESO coefficient value of the unidimensional logical thinking ability test items is lower when compared to the multidimensional ESO coefficient value due to the validity testing of the unidimensional logical thinking ability test items which was carried out before and after being corrected for the effects *spurious overlap* involves a higher number of items than multidimensional testing. This is very different from testing the validity of multidimensional logical thinking ability test items which is carried out before and after being corrected for the spurious overlap effect. Testing the validity of test items for logical thinking abilities in the dimensions of deductive thinking and inductive thinking involves fewer test items, so the spurious overlap effect is relatively large. So the difference between the rix coefficient and $r_i(x_i)$ is multidimensionally larger. From this section, it can be said that in unidimensional testing of logical thinking ability test items, which was carried out before and after being corrected for the spurious overlap effect, the contribution of the spurious overlap effect was smaller when compared to those tested multidimensionally. However, an ESO value that is greater than a comparison can be interpreted as a more careful process of testing the validity of an instrument's items. This is because the items whose validity is tested come from

the same dimension. From this statement it can be suggested that if you want to test the validity of the items of an instrument carefully you should test the validity of the items according to their dimensions.

5. REFERENCES

- Abate, T., Michael, K., & Angell, C. (2020). Assessment of scientific reasoning: Development and validation of scientific reasoning assessment tool. *EURASIA Journal of Mathematics, Science and Technology Education*, 16(12). <https://doi.org/10.29333/ejmste/9353>.
- Agung, A. A. G. (2018). *Metodologi Penelitian Kuantitatif (Perspektif Manajemen Pendidikan)*. Universitas Pendidikan Ganesha.
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/10.21831/pep.v22i2.19590>.
- Baskara, G. W., & Yudiana, K. (2021). Tes Tertulis Berbasis HOTS pada Pembelajaran IPA Menggunakan Google Form Siswa Kelas IV SD. *Mimbar PGSD Undiksha*, 9(3), 416–423. <https://doi.org/10.23887/jjpsd.v9i3.39223>.
- Bhaw, N., Kriek, J., & Lemmer, M. (2023). Insights from coherence in students' scientific reasoning skills. *Heliyon*, 9(7), e17349. <https://doi.org/10.1016/j.heliyon.2023.e17349>.
- Darmayasa, I. K., Jampel, N., & Simamora, A. (2018). Pengembangan E-Modul Ipa Berorientasi Pendidikan Karakter Di Smp Negeri 1 Singaraja. *Jurnal Edutech*, 6(1), 53–65. <https://doi.org/10.23887/jeu.v6i1.20267>.
- Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Analisis kualitas soal kemampuan membedakan rangkaian seri dan paralel melalui teori tes klasik dan model rasch. *Indonesian Journal of Educational Research and Review*, 3(1), 11–19. <https://doi.org/10.23887/ijerr.v3i1.24080>.
- Janna, N. M., & Herianto. (2021). Konsep uji validitas dan reliabilitas. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, 18210047, 1–12.
- Karjo, Ashadi, & Sugiyarto. (2019). Penerapan Pendekatan Saintifik Dalam Pembelajaran Biologi Menggunakan Model GI Dan STAD Ditinjau Dari Kemampuan Berpikir Deduktif Dan Motivasi Berprestasi. *Inkuiri: Jurnal Pendidikan IPA*, 8(2), 162 – 171. <https://doi.org/10.20961/inkuiri.v8i2.37755>.
- Khasanah, M. M., Ramli, M., & Dwiastuti, S. (2020). Developing a Dynamic Assessment Instrument to Assess Reasoning Skills About Bacteria. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 24(1), 62–75. <https://doi.org/10.21831/pep.v24i1.28815>.
- Kurnia, A., Sukarmin, & Sunarno, W. (2021). Profil Kemampuan Berpikir Kreatif Siswa Menggunakan Soal Tes Pilihan Ganda pada Pembelajaran Ilmu Pengetahuan Alam. *Indonesian Journal of Educational Science (IJES)*, 4(1), 27–32. <https://doi.org/10.31605/ijes.v4i1.1147>.
- Larasati, D. A., & Widyastuti, T. (2020). Pengembangan Skala Pemaafan Diri (Self-forgiveness). *Acta Psychologia*, 2(1), 80–90. <https://doi.org/10.21831/ap.v1i1.34121>.
- Leonard, K., Sepehri, P., Cheri, B., & Kelly, D. M. (2023). Relational complexity influences analogical reasoning ability. *Iscience*, 26(4), 106392. <https://doi.org/10.1016/j.isci.2023.106392>.
- Lestari, N., & Harjono, N. (2021). Pengembangan Instrumen Penilaian Karakter PPK Aspek Kemandirian Pada Siswa SD Kelas 4. *MIMBAR PGSD Undiksha*, 9(1). <https://doi.org/10.23887/jjpsd.v9i1.33379>.
- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV Mi Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86–96. <https://doi.org/10.23887/jisd.v4i1.23335>.
- Nuraeni, M. I., & Habibi, M. W. (2021). Pengembangan Bahan Ajar IPA Berbentuk Komik Pada Materi Sistem Ekskresi Untuk SMP/MT's Kelas VIII. *Experiment: Journal of Science Education*, 1(1). <https://doi.org/10.18860/experiment.v1i1.11116>.
- Paramartha, W. E., Dharsana, I. K., & Padmi, N. M. D. (2022). Kalibrasi Item Instrumen Kecenderungan Nomophobia Pada Remaja Milenial. *Jurnal Bimbingan Dan Konseling Indonesia*, 7(2), 339–349. https://doi.org/10.23887/jurnal_bk.v7i2.1581.
- Pelikan, A., Herzel, H., Kramer, A., & Ananthasubramaniam, B. (2022). Venn diagram analysis overestimates the extent of circadian rhythm reprogramming. *The FEBS Journal*, 289(21), 6605–6621. <https://doi.org/10.1111/febs.16095>.
- Puger, I. G. N., & Dewi, K. Y. F. (2023). Validitas Butir Kuesioner Perilaku Siswa Berbasis Ajaran Karmapatha Secara Unidimensi Dan Multidimensi Sebelum Dikoreksi Oleh Efek Spurious Overlap. *Daiwi Widya*, 10(1), 1–12. <https://doi.org/10.37637/dw.v10i1.1387>.
- Purnama Dewi, L. P. (2018). Penerapan Model Pembelajaran Quantum Teaching berbantuan Media Konkret untuk Meningkatkan Hasil Belajar IPA. *Journal of Education Action Research*, 2(1), 23.

- <https://doi.org/10.23887/jear.v2i1.13727>.
- Putra, A., Syarifuddin, H., & Zulfah, Z. (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. <https://doi.org/10.32939/ejrpm.v1i2.302>.
- Rezeki, S., & Ishafit, I. (2017). Pengembangan Media Pembelajaran Interaktif untuk Sekolah Menengah Atas Kelas XI pada Pokok Bahasan Momentum. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 3(1), 29. <https://doi.org/10.21009/1.03104>.
- Septriningsih, A. S. ., Bektiarso, S., & T, P. (2010). Pengaruh Model Problem Based Instruction Terhadap Keterampilan Proses Sains dan Hasil Belajar Siswa Dalam Pembelajaran IPA di SMP. *Jurnal Pembelajaran IPA*, 3(6).
- Talman, K., Vierula, J., Kanerva, A. M., Virkki, O., Koivisto, J. M., & Haavisto, E. (2021). Instruments for assessing reasoning skills in higher education: A scoping review. *Assessment & Evaluation in Higher Education*, 46(3), 376–392. <https://doi.org/10.1080/02602938.2020.1776212>.
- Voskoglou, M., & Broumi, S. (2022). A Hybrid Method For The Assessment of Analogical Reasoning Skills. *Journal of Fuzzy Extension and Applications*, 3(2), 152–157. <https://doi.org/10.22105/jfea.2022.342139.1220>.
- Warju, W., Ariyanto, S. R., Soeryanto, S., & Trisna, R. A. (2020). Analisis Kualitas Butir Soal Tipe HOTS pada Kompetensi Sistem REM di Sekolah Menengah Kejuruan. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 17(1), 95–104. <https://doi.org/10.23887/jptk-undiksha.v17i1.22914>.
- Wulandari, A. I., & Radia, E. H. (2021). Pengembangan Instrumen Penilaian Sikap Tanggung Jawab Pembelajaran Tematik Terpadu Kelas V SD. *MIMBAR PGSD Undiksha*, 9(1), 10–18. <https://doi.org/10.23887/jjgsd.v9i1.32979>.
- Xu, W., Geng, F., & Wang, L. (2022). Relations of computational thinking to reasoning ability and creative thinking in young children: Mediating role of arithmetic fluency. *Thinking Skills and Creativity*, 4, 101041. <https://doi.org/10.1016/j.tsc.2022.101041>.
- Yugakisha, M. S., & Jayanta, I. N. L. (2021). Instrumen Penilaian Motivasi Belajar Ipa. *Jurnal Pedagogi Dan Pembelajaran*, 4(2), 402–408. <https://doi.org/10.23887/jp2.v4i2.35918>.
- Yusup, F. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23. <https://doi.org/10.21831/jorpres.v13i1.12884>.