International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

October 2022 ● Vol.15, No.4
p-ISSN: 1694-609X
pp. 535-554

Article submission code: 20211023091122

Received: 23/10/2021
Revision: 10/05/2022

Accepted: 29/05/2022
OnlineFirst: 06/08/2022

Exploratory Analysis on Adaptive Reasoning of Undergraduate Student in Statistical Inference

Karunia Eka Lestari
Corresponding author, Faculty of Teacher Training and Education, Universitas Singaperbangsa Karawang, Indonesia, *karunia@fkip.unsika.ac.id*

Marsah Rahmawati Utami
Faculty of Health Science, Universitas Singaperbangsa Karawang Indonesia, *marsah.rahmawati@fikes.unsika.ac.id*

Mokhammad Ridwan Yudhanegara
Faculty of Teacher Training and Education, Universitas Singaperbangsa Karawang Indonesia, *mridwan.yudhanegara@staff.unsika.ac.id*

In statistical inference, adaptive reasoning is defined as logical thinking to determine what can be inferred from data or statistical results and whether the justifications led to valid conclusions. Accordingly, adaptive reasoning is a mathematical proficiency required in statistical inference. This study aims to discover the association between adaptive reasoning and the initial statistical competence of undergraduate students. For this purpose, we performed mixed-methods research conducted by sequential exploratory analysis. This study involved 66 participants selected from undergraduate students in the Statistical Inference course offered by the mathematics education department at one university in Indonesia. The qualitative result describes the characteristics of students' adaptive reasoning proficiency at each grade. The proportion of students from grade 1 to grade 4 is 4.55%, 21.21%, 48.48%, and 25.76%, respectively. The quantitative result based on the chi-squared statistics test shows a significant association between adaptive reasoning proficiency and initial statistical competence. The correspondence analysis solution depicts that a high level of statistical competence is strongly associated with a high grade of adaptive reasoning proficiency, and conversely. Generally, the results provide evidence that the mastery of initial statistical competence is an important aspect in developing students' adaptive reasoning proficiency. The study provides some recommendations that will benefit the lecturer to develop adaptive reasoning proficiency in the Statistical Inference courses.

Keywords: education statistics, exploratory analysis, correspondence analysis, adaptive reasoning, k-means clustering

Citation: Lestari, K. E., Utami, M. R., & Yudhanegara, M. R. (2022). Exploratory analysis on adaptive reasoning of undergraduate student in statistical inference. *International Journal of Instruction*, *15*(4), 535-554. https://doi.org/10.29333/iji.2022.15429a

INTRODUCTION

Statistics education research has drastically shifted over the past few years from focusing on procedures-calculating a mean, interpreting box plots, comparing groups-towards a greater focus on statistical reasoning. Statistical reasoning is the way people reason with statistical ideas and make sense of statistical information, leading to inference and interpreting statistical results (Garfield, 2002). From the inferential statistics point of view, statistical reasoning involves an understanding of how used the methods to estimate or predict unknown characteristics of interest by using available data or how samples are related to a population and what may be inferred from a sample (Garfield, 2003; Lesik, 2019).

Research in statistics education has long suggested that students have difficulty using statistical reasoning appropriately in applied problems (Makar & Rubin, 2009). For example, research on elementary statistics courses at university suggests that even students who could successfully implement hypothesis testing and parameter estimation procedures could not use these procedures appropriately in applications (Gardner & Hudson, 1999; Reichardt & Gollob, 1997). Erickson (2006) argued that inference is so complex that even professional researchers misuse it.

However, the material object in statistical inference is often complex for students and teachers to understand (Lugo-Armenta & Pino-Fan, 2021). Several studies have identified errors and difficulties presented by both when making inferences, especially on the understanding of the significance level, type I and type II error, the logic of hypothesis testing, the formulation of statistical hypotheses and sampling distributions, and the relationship between the statistic and the parameter (Sotos et al., 2007; Garfield & Ben-Zvi, 2008; Harradine et al., 2011; Hupe, 2015).

Since statistical inference relies on generalizing to a larger population, it is essential to understand what that population is. The backbone of making inferences about a population parameter depends on using information obtained from a sample statistic. For instance, we can make inferences about an unknown population average using the sample average. Once we understand the distribution pattern of the sample statistics, we can then use these sample statistics to calculate confidence intervals and perform hypothesis tests about an unknown population parameter. In addition to making inferences about population average, we can also infer population proportions, population variances, and other population parameters.

Statistical reasoning requires logical thinking and the ability to reason and justify why solutions are appropriate. This study defined statistical reasoning in inferential statistics as statistical competence. Specifically, this competence includes understanding the data, hypothesizing what the data might show, investigating the hypothesis, and writing a valid conclusion. These abilities refer to adaptive reasoning (Ostler, 2011; Nurjanah et al., 2021).

Adaptive reasoning is a strand in mathematical proficiency that refers to thinking logically about the relationships among concepts and situations, including proficiency in reflection, explanation, and justification (Kilpatrick et al., 2001). Adaptive reasoning is viewed as an activity needed in mathematics because mathematics is an abstract science or an abstract domain that requires high-level reasoning to understand it (Staub & Stern, 1997; Risnawati et al., 2019). In statistical inference, adaptive reasoning is defined as logical thinking to determine what can be inferred from data or statistical results and whether the justifications led to valid conclusions (Reid & Cox, 2015). One can use it to glue many facts, procedures, concepts, and solution methods and to see that they all fit together in some way, that they make sense. Therefore, adaptive reasoning proficiency is measured based on four indicators: reflection abilities, logical thought, explanation, and justification.

In the last five years, research on adaptive reasoning proficiency commonly analyzed descriptively or through correlational relationships to other mathematical skills (Goldstone et al., 2017; Awofala, 2017; McFeetors & Palfy, 2018; Lepak et al., 2018; Max & Welder, 2020; Víquez, 2022). Instead, this study emphasizes graphical analysis of the relationship between adaptive reasoning ability and initial statistical competence using correspondence analysis by first creating categorical variables through k-means clustering. This analysis is a novelty that has not been found in any paper, particularly in educational research. By considering the importance of adaptive reasoning to statistical competence, this study explores the adaptive reasoning of undergraduate students by (1) describing the adaptive reasoning based on their level of initial statistical competence; (2) discovering the association (dependency) structure between adaptive reasoning and initial statistical competence. As a limitation in this study, the association refers to the dependence between variables which is measured by Pearson's chi-square statistic and analyzed using correspondence analysis.

METHOD

Research Design

The mixed-methods research is conducted by sequential exploratory analysis. Creswell (2018) explained that exploratory analysis in mixed-methods research involves qualitative data collection and analysis in the first stage, followed by the second stage of quantitative data collection and analysis that builds on the result of the first stage. Exploratory research is conducted in order to obtain new perceptions, gain new ideas, and widen the knowledge of a phenomenon (Schutt, 2014; Hembrough & Jordan, 2020).

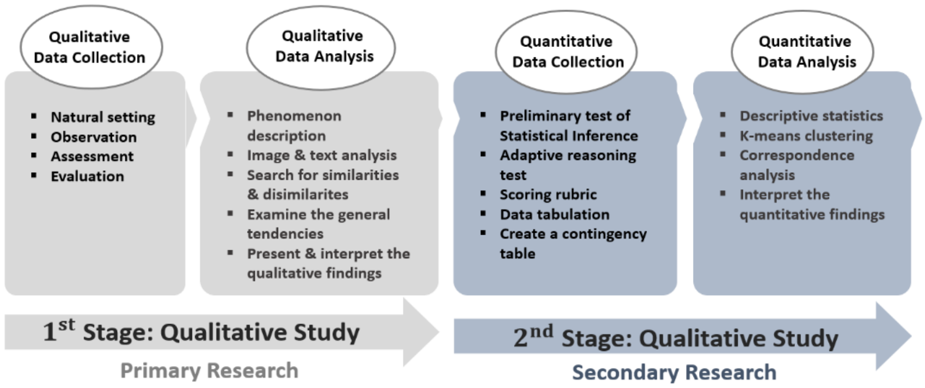


Figure 1
Sequential exploratory analysis design

The primary priority in this design emphasizes the collection and analysis of qualitative data and then explains the quantitative result with in-depth qualitative data. The study begins with qualitative research that is conducted by grounded theory methods. In this method, the students are observed in a natural setting to discover a theory grounded in information from students. Then the students were given a preliminary test to determine initial statistical competence. At the end of the first stage, students' adaptive reasoning assessment and evaluation are executed. The results are presented in descriptive. From this initial exploration, the qualitative findings were used to examine the association between adaptive reasoning and initial statistical competence. In other words, the qualitative stage will be conducted as a follow up to the quantitative results to help explain the quantitative results. Furthermore, quantitative data analysis was carried out using k-means clustering and correspondence analysis. The design in this study is illustrated as in Figure 1.

Participant

This study involved 66 participants selected from 163 undergraduate students in the Statistical Inference course offered by the mathematics education department at one university in Indonesia. The sample is established through the purposive sampling technique.

Data Collection

There are two schemes of data collection that are collected sequentially. For the initial, the qualitative data collection involves observation, assessment and evaluation. The researcher has firsthand experience with participants, records information as it occurs, and notices the unusual phenomenon during observation. Next, the quantitative data collection involves testing, scoring rubric, tabulating data, and creating a contingency table. There are two types of tests, including a preliminary test for statistical competence and the adaptive reasoning test. The preliminary test measured students' ability to understand data, formulate and test the hypothesis, write a valid conclusion, and interpret statistical results. Meanwhile, adaptive reasoning proficiency is measured based on four indicators: reflection, logical thought, explanation, and justification.

Data Analysis

The qualitative data were analyzed descriptively using text and image analysis. The data described are students' answers on the adaptive reasoning test. This analysis was carried out through four stages, referring to Lamnek (2005):

1. Search for similarities among all or some of the students' responses.
2. Determine differences in the content of the student's responses.
3. Examine the similarities and differences to determine general tendencies.
4. Present and interpret various types of participants, statements, information, etc., considering the individual cases.

Quantitative data were analyzed using k-means clustering and correspondence analysis. K-mean clustering was performed to find students' similarities based on their ISC and classify them into several groups (Yudhanegara et al., 2021). In this clustering, the number of clusters is determined at the beginning, then the researcher identifies and interprets the clusters formed subjectively (Yudhanegara et al., 2020a, 2020b). Specifically, k-means clustering was used in this study to divide the student into three levels based on their ISC. The determination of the three clusters considers distributing ISC scores that expect approximately to a normal distribution, with the proportion of 50% of students being at the average level, the proportion of excellent and below-average levels being 25% each (Yudhanegara & Lestari, 2019).

FINDINGS

Data descriptions

The descriptive statistics of the preliminary test for statistical competence and adaptive reasoning test are summarized in Table 1. In these results, the average score of initial statistical competence (ISC) is 69.29, and the median score is 73. Since the average is greater than the median, thus the data appear to be skewed to the right. It indicates that most ISC scores accumulate in lower value (left side) and few in the higher value. Conversely, the average score for adaptive reasoning proficiency (ARP) is 72.88, and the median is greater than the average. Thus, the data score of adaptive reasoning appear to be skewed to the left, or the most of scores are in high values (right side) and a few in the lower value (right side).

Table 1
Descriptive statistics that describe the characteristics of students' abilities in Statistical Inference course

Descriptive statistics	Initial Statistical Competence	Adaptive Reasoning Proficiency
Average	69.29	72.88
Standard deviation	12.55	10.43
Minimum	36	33
Fist quartile	65.25	66.25
Median	73	74
Third quartile	76	78.75
Maximum	93	99
Range	57	66

Often, skewness is easiest to detect by histograms, as presented in Figure 2. By these histograms, we could examine the shape of data to determine whether the data appear to be skewed.

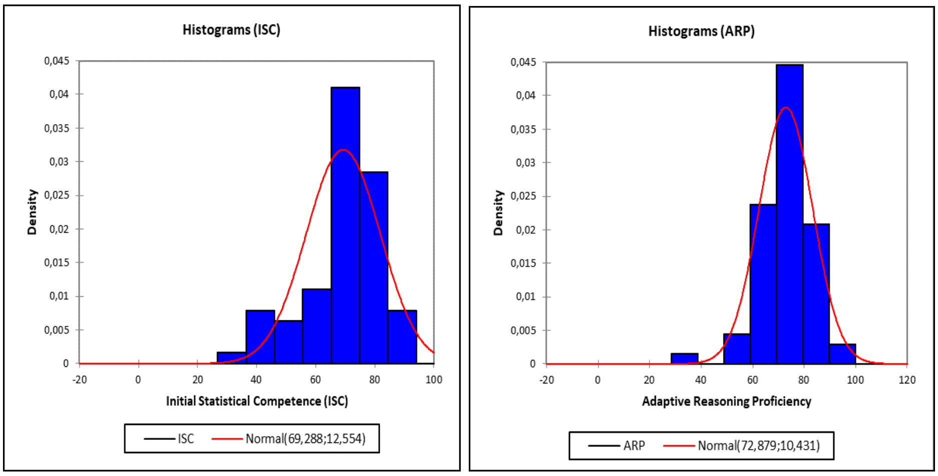


Figure 2
Histogram of the score on a preliminary test of Statistical Inference (left), and histogram of the score on adaptive reasoning test (right)

According to Table 1, the standard deviation for ISC and ARP are 12.55 and 10.43, respectively. Thus the standard deviation for ISC is greater than the standard deviation for ARP. A higher standard deviation value indicates a greater spread in the data. It implies that ARP scores are more spread out from its centre than the ISC scores, as captured in Figure 3.

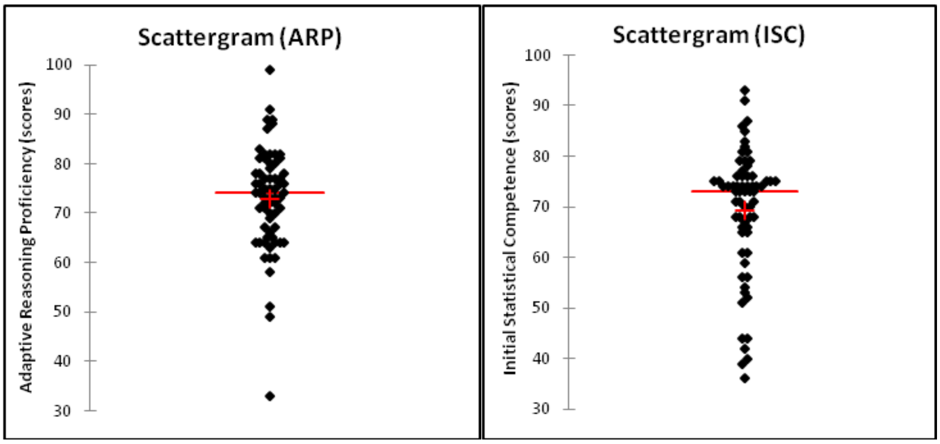


Figure 3
Scattergram of the score on a preliminary test of Statistical Inference (left), and scattergram of the score on adaptive reasoning test (right)

Overall, a visual summary of the data's shape, spread, and centre is displayed in the following boxplots. These plots show the five-number summary of a set of data, including the minimum, first quartile, median, third quartile, and maximum scores.

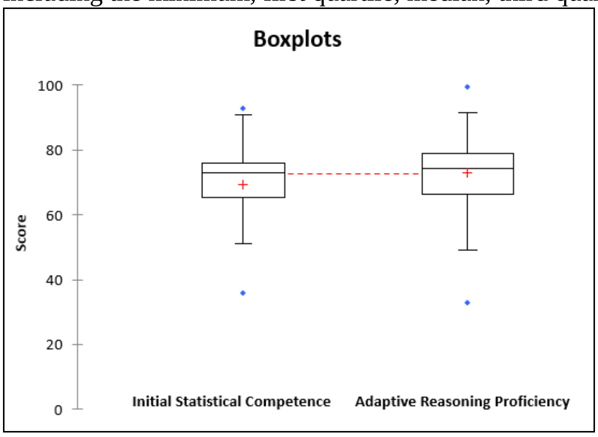


Figure 4
Boxplots of the score on a preliminary test of Statistical Inference and score on the adaptive reasoning test

In the boxplots above, the dashed-red line that connected two boxplots referred to the median line. This line lies inside of the box of a comparison boxplot. It indicates no differences between scores on a preliminary test and scores on the adaptive reasoning test. In addition, the two boxplots contain outliers, both major outliers and minor outliers. It shows that there are students whose scores are far away from the scores of other students.

Furthermore, the statistical result of k-means clustering for scores on a preliminary test of Statistical Inference is recorded in Table 2.

Table 2
K-means clustering for scores on a preliminary test

Descriptive statistics	Cluster 1	Cluster 2	Cluster 3
Cluster member	13	13	40
Within cluster-average (centroid)	48.15	83.38	71.58
Within-clas variance	57.64	22.76	17.43
Minimum distance to centroid	2.85	0.38	0.57
Average distance to centroid	6.76	3.86	3.46
Maximum distance to centroid	12.15	9.62	10.58

As noted before, the clustering results are identified by their centroid or within-cluster average and subjectively interpreted. By comparing the centroids of each cluster and rearranging them in ascending order, it is found that the average of cluster 3 is more than the average of cluster 1 but less than the average of cluster 2. Finally, we define cluster 1 as "below-average", cluster 3 as "average", and cluster 2 as "excellent". The membership list of each cluster is summarized in Table 3.

Table 3
Cluster membership is based on the initial statistical competence (in ascending order); the centre of each cluster is marked in bold

Below-average	Average	Excellent
S1, S4, S5, S7, S14, S15, S17, S20, S23, S24, S25, S28, S33	S8, S10, S11, S12, S13, S16, S18, S19, S21, S22, S30, S31, S32, S36, S37, S38, S39, S40, S41, S42, S43, S44, S45, S47, S48, S50, S51, S52, S53, S54, S55, S56, S57, S58, S60, S61, S62, S64, S65, S66	S2, S3, S6, S9, S26, S27, S29, S34, S35, S46, S49, S59, S63

Fascinatingly, the clustering results with k-means clustering obtained the proportion below-average: average: excellent as expected, which is 0.20:0.60:0.20. Thus, the initial statistical competence appears roughly normally distributed. As a note, since this is an exploratory analysis, we should not expect to see a perfectly normal distribution.

Another quantitative data was obtained from the adaptive reasoning test. This data was analyzed using correspondence analysis to discover the association structure between adaptive reasoning and initial statistical competence. Correspondence analysis (CA) is a popular graphical method used to explore the association (dependency) structure between categorical variables in a contingency table (Beh et al., 2010; Lestari et al., 2019a). In addition, CA is a graphical method for representing information of categorical variables in a contingency table (Rencer & Christensen, 2012; Lestari et al., 2019b).

Based on these references, two main aspects in CA are (1) categorical variables and (2) contingency tables. To reach these two aspects, we classified adaptive reasoning proficiency into four grades arranged in ascending order, where each grade reflects the number of indicators that students achieve. This data is then cross-classified with initial statistical competence data obtained from previously clustering. From here, we have two categorical variables, both of which have an ordinal scale. Initial statistical competence as a row-categorical variable consists of three categories, including "below-average", "average", and "excellent". Adaptive reasoning as a column-categorical variable consists of four categories, including "grade 1", "grade 2", "grade 3", and "grade 4". The results of the cross-classification of the two variables are tabulated in a table known as a contingency table (Lestari et al., 2020) and visualized as follows.

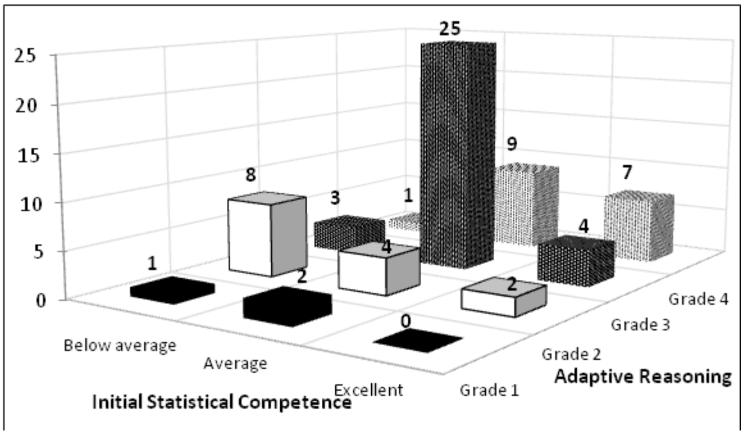


Figure 5
The three-dimensional view of the contingency table that cross-classified 66 undergraduate students according to initial statistical competence and their adaptive reasoning proficiency

The figure above shows that the highest frequency is the students classified at the average statistical competence level and adaptive reasoning at grade 3, with a percentage of 37.88%. At the same time, none of the students at the excellent level was in grade 1.

Qualitative Findings

The qualitative study in the first stage aims to describe the adaptive reasoning based on their level of statistical competence. The qualitative findings are obtained from students' answers on the adaptive reasoning test. This test consists of five questions that measure students' adaptive reasoning on statistical inference for one sample, the difference between two means, analysis of variance, linear regression, and nonparametric statistics. Each question measures four indicators of adaptive reasoning proficiency, including reflection, logical thought, explanation, and justification. One of five questions will be analyzed by description and displayed in Figure 6.

Pada masa pandemi COVID-19, seorang dosen Prodi Pendidikan Matematika melakukan penelitian dengan menerapkan pembelajaran campuran (*blended*) pada perkuliahan Statistika Inferensial. Kemudian dia mengambil sampel acak dari 10 siswa yang memperoleh pembelajaran tersebut, dengan rataan dan simpangan baku sampel berturut-turut ialah $\bar{x} = 80{,}2$ dan $s = 10{,}91$. Dosen tersebut mengklaim bahwa rata-rata nilai UAS mahasiswa yang memperoleh pembelajaran campuran lebih dari rata-rata tahun sebelumnya. Sementara itu, berdasarkan data dari tahun sebelumnya, diketahui bahwa rata-rata nilai UAS mahasiswa pada mata kuliah Statistika Inferensial adalah 82. Jika sampel yang dipilih diasumsikan berasal dari populasi yang **berdistribusi normal**, tentukan: a. **[Refleksi]** Estimasi selang kepercayaan 95% yang menunjukkan bahwa parameter rataan populasi μ akan berada pada interval $\left(\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$. b. **[Berpikir Logis]** Tentukan uji statistik untuk menguji klaim tersebut dengan taraf signifikansi tertentu. c. **[Eksplanasi]** Berikan interpretasi terkait kesimpulan yang diperoleh dari hasil pengujian yang dilakukan. d. **[Justifikasi]** Berdasarkan kesimpulan pada poin c, apakah perlu dilakukan uji kuasa?	During the COVID-19 pandemic, a lecturer in the Mathematics Education Department applied blended learning methods in the Inferential Statistics course. A random sample of 10 students who obtained the method was selected, with the mean and standard deviation being $\bar{x} = 80{,}2$ and $s = 10{,}91$, respectively. The lecturer claims that the average score of the final exam for students who receive blended learning is more than the average score from the previous year's average. Historical data of the last year shows that the average score of the final exam is 82. Assuming that the population where are sampling from is **normally distributed**, determine: a. **[Reflection]** Estimated 95% confidence interval for the unknown population mean μ score of the final exam will be in $\left(\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$. b. **[Logical thought]** Test the lecturer's claim for a certain significance level. c. **[Explanation]** Interpret the statistical result. d. **[Justification]** Based on the result in point c, wheater to perform a power analysis?

Figure 6
The 4th question of adaptive reasoning test on statistical inference for one sample T-test; the original question (left) and in English version (right)

For the first step, we search for the similarity of students' answers to these questions. Students who can answer one indicator are classified in "grade 1", those who answer two indicators in "grade 2", and so on until "grade 4". In the second step, we determine the difference in content for each grade. The 4th question is chosen to be examined by considering the similarities and dissimilarities among all responses. The similarities and differences found in the previous two steps were used to determine general tendencies regarding students' adaptive reasoning. The results show that almost half of the students (reached 48.48%) can achieve the three indicators and are classified as "grade 3", and only three students or about 4.55% were in grade 1. The proportions for each grade as a whole are displayed in Figure 7.

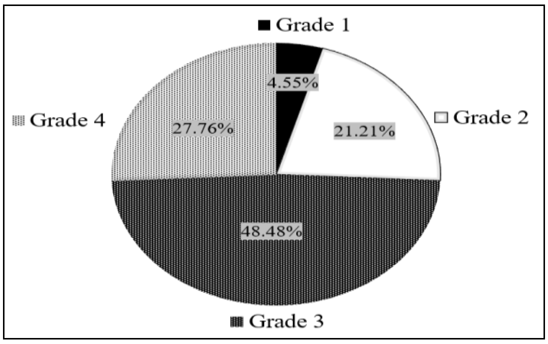
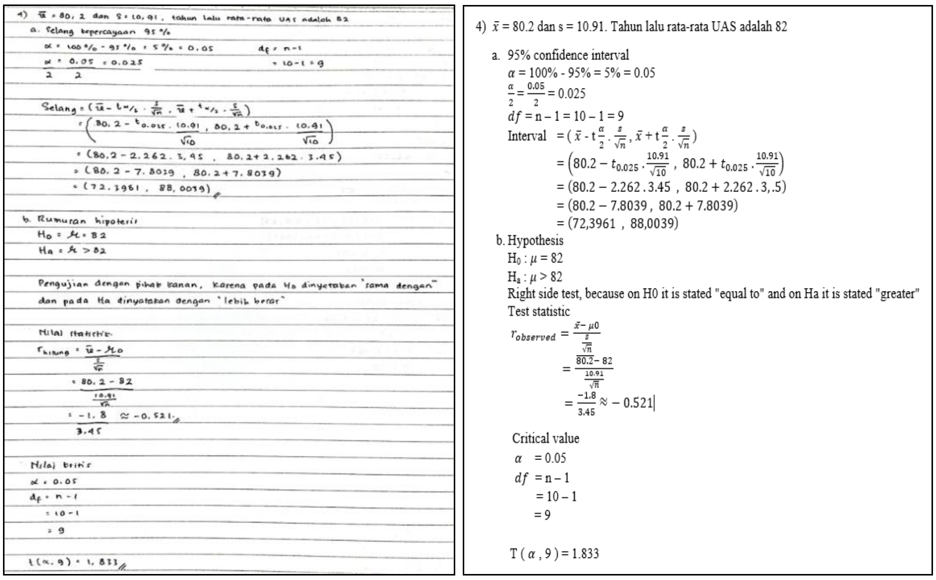


Figure 7
The proportion of students based on adaptive reasoning grade

In the final stage, we present and interpret the various characteristics of participants, statements, and information based on the preferences in each grade. Figure 8 displays the answer from one of the three students in grade 1. Students in grade 1 can estimate the confidence interval for the unknown population average μ at a certain α-level of significance. In point (a), the formula to determine the confidence interval is given; students are only required to reflect on what is known in the problem and then apply this formula. But when asked to test the hypothesis, students cannot do it thoroughly. The student can formulate hypotheses, determine statistics and critical values but can not examine the criterion and make a valid conclusion. In other words, students in grade 1 could not get logical thought, construct explanation and justification.



4) $\bar{x}$ = 80, 2 dan s = 10, 91, tahun lalu rata-rata UAS adalah 82

a. Selang kepercayaan 95 %

α = 100 % - 95 % = 5 % = 0, 05 $\quad$ $d_f = n - 1 = 10 - 1 = 9$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

$$\text{Selang} = \left(\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$
$$= \left(80.2 - t_{0.025} \cdot \frac{10.91}{\sqrt{10}},\ 80.2 + t_{0.025} \cdot \frac{10.91}{\sqrt{10}}\right)$$
$$= (80.2 - 2.262 \cdot 3,45\ ,\ 80.2 + 2.262 \cdot 3.45)$$
$$= (80.2 - 7.8039\ ,\ 80.2 + 7.8039)$$
$$= (72.3961\ ,\ 88,0039)$$

b. Rumusan hipotesis

$H_0 = \mu = 82$

$H_a = \mu > 82$

Pengujian dengan pihak kanan, karena pada Ho dinyatakan "sama dengan" dan pada Ha dinyatakan dengan "lebih besar"

Nilai statistik

$$t_{hitung} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{80.2 - 82}{\frac{10.91}{\sqrt{n}}} = \frac{-1.8}{3.45} \approx -0.521$$

Nilai kritis

$\alpha = 0.05$

$d_f = n - 1 = 10 - 1 = 9$

$t(\alpha, 9) = 1.833$

4) $\bar{x}$ = 80.2 dan s = 10.91. Tahun lalu rata-rata UAS adalah 82

a. 95% confidence interval

α = 100% - 95% = 5% = 0.05

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

df = n – 1 = 10 – 1 = 9

$$\text{Interval} = \left(\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right)$$
$$= \left(80.2 - t_{0.025} \cdot \frac{10.91}{\sqrt{10}}\ ,\ 80.2 + t_{0.025} \cdot \frac{10.91}{\sqrt{10}}\right)$$
$$= (80.2 - 2.262\,.3.45\ ,\ 80.2 + 2.262\,.3,.5)$$
$$= (80.2 - 7.8039\ ,\ 80.2 + 7.8039)$$
$$= (72{,}3961\ ,\ 88{,}0039)$$

b. Hypothesis

$H_0 : \mu = 82$

$H_a : \mu > 82$

Right side test, because on H0 it is stated "equal to" and on Ha it is stated "greater"

Test statistic

$$r_{observed} = \frac{\bar{x} - \mu 0}{\frac{s}{\sqrt{n}}} = \frac{80.2 - 82}{\frac{10.91}{\sqrt{n}}} = \frac{-1.8}{3.45} \approx -0.521$$

Critical value

α = 0.05

df = n – 1 = 10 – 1 = 9

T (α , 9) = 1.833

Figure 8
Students' answers in grade 1 on the adaptive reasoning test; the original response (left) and in English version (right)

Figure 9 presents the answer from one of the 14 students in grade 2. Students in grade 2 could estimate the unknown population's confidence interval average μ at a certain α-level of significance, formulating, investigating, and testing the hypothesis. But when asked to interpret the statistical result, students could not explain it until getting a valid conclusion. It indicated that students in grade 2 could not construct the explanation and justification.

On the other hand, Figure 10 displays the response from one of the 32 students in grade 3. Students in grade 3 could estimate the unknown population's confidence interval average μ at a certain α-level of significance, perform hypothetical testing, interpret a statistical result, and write a valid conclusion. But when asked whether or not to perform a power analysis based on the previous testing, students commonly could not justify it. In other words, students in grade 3 could not examine a justification according to previous solutions.

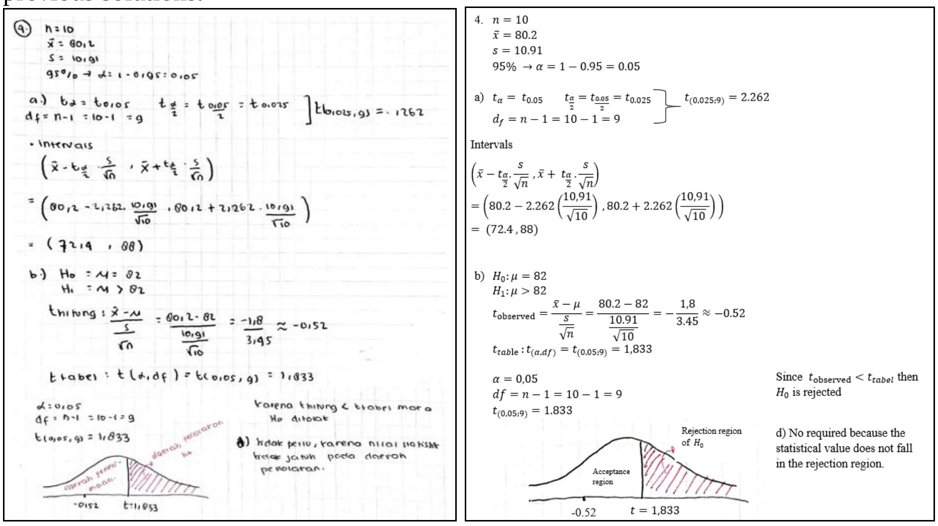


Figure 9

Students' answers in grade 2 on the adaptive reasoning test; the original response (left) and in English version (right)

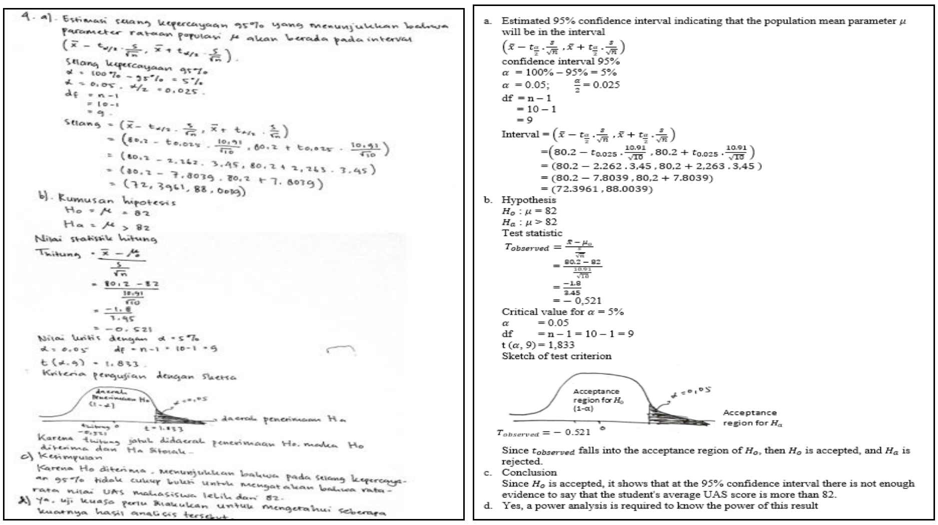
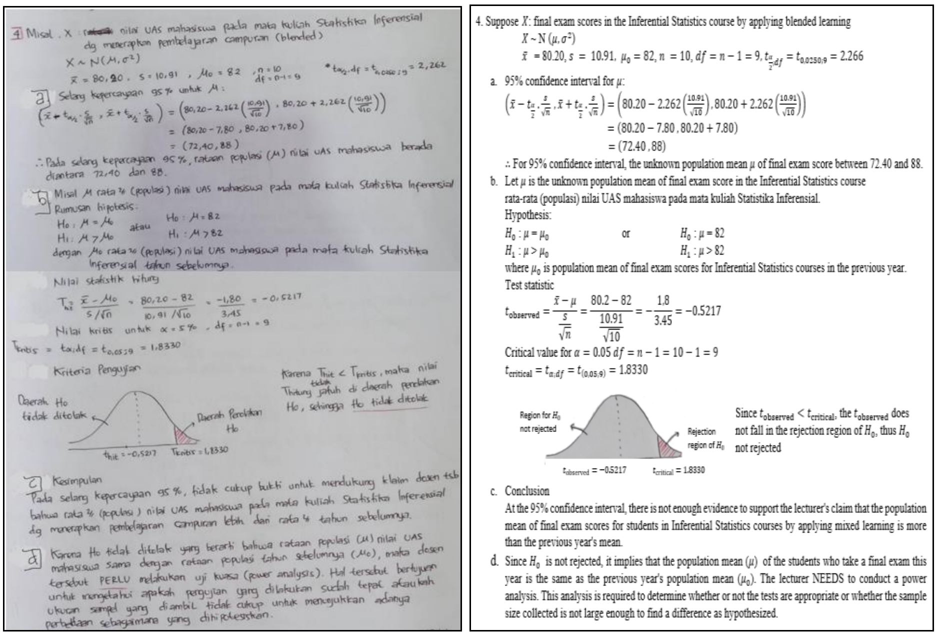


Figure 10

Students' answers in grade 3 on the adaptive reasoning test; the original response (left) and in English version (right)

In comparison, students in grade 4 could complete the four indicators of adaptive reasoning proficiency, such as estimating the unknown population's confidence interval average μ at a certain α-level of significance, performing hypothetical testing, interpreting a statistical result, writing a valid conclusion, and examining a justification according to previous solutions. The response from one of the 17 students in grade 4 is presented in Figure 11 below.


4. Suppose X: final exam scores in the Inferential Statistics course by applying blended learning

$X \sim N(\mu, \sigma^2)$

$\bar{x} = 80.20, s = 10.91, \mu_0 = 82, n = 10, df = n-1 = 9, t_{\frac{\alpha}{2};df} = t_{0{,}0250;9} = 2.266$

a. 95% confidence interval for μ:

$$\left(\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right) = \left(80.20 - 2.262\left(\frac{10.91}{\sqrt{10}}\right), 80.20 + 2.262\left(\frac{10.91}{\sqrt{10}}\right)\right)$$
$$= (80.20 - 7.80, 80.20 + 7.80)$$
$$= (72.40, 88)$$

∴ For 95% confidence interval, the unknown population mean μ of final exam score between 72.40 and 88.

b. Let μ is the unknown population mean of final exam score in the Inferential Statistics course rata-rata (populasi) nilai UAS mahasiswa pada mata kuliah Statistika Inferensial.

Hypothesis:

$H_0 : \mu = \mu_0$ or $H_0 : \mu = 82$

$H_1 : \mu > \mu_0$ $H_1 : \mu > 82$

where μ_0 is population mean of final exam scores for Inferential Statistics courses in the previous year.

Test statistic

$$t_{\text{observed}} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{80.2 - 82}{\frac{10.91}{\sqrt{10}}} = -\frac{1.8}{3.45} = -0.5217$$

Critical value for $\alpha = 0.05$ $df = n - 1 = 10 - 1 = 9$

$t_{\text{critical}} = t_{\alpha;df} = t_{(0{,}05;9)} = 1.8330$

Since $t_{\text{observed}} < t_{\text{critical}}$, the t_{observed} does not fall in the rejection region of H_0, thus H_0 not rejected

c. Conclusion

At the 95% confidence interval, there is not enough evidence to support the lecturer's claim that the population mean of final exam scores for students in Inferential Statistics courses by applying mixed learning is more than the previous year's mean.

d. Since H_0 is not rejected, it implies that the population mean (μ) of the students who take a final exam this year is the same as the previous year's population mean (μ_0). The lecturer NEEDS to conduct a power analysis. This analysis is required to determine whether or not the tests are appropriate or whether the sample size collected is not large enough to find a difference as hypothesized.

Figure 11

Students' answers in grade 4 on the adaptive reasoning test; the original response (left) and in English version (right)

Quantitative Findings

The quantitative study in the second stage aims to discover the association (dependency) structure between adaptive reasoning and initial statistical competence. This association is visualized in Figure 12. The figure shows that it appears no linear relationship exists between the two variables. But it does not suggest that no association exists among variables. To explore this association, we conducted a correspondence analysis. The association between two categorical variables was tested using Pearson's chi-squared χ^2 statistic with hypotheses as follows:

H_0: Initial statistical competence and adaptive reasoning are **not** associated

H_1: Initial statistical competence and adaptive reasoning are associated.

If $\chi^2 > \chi^2_{\alpha,v}$ with $v = (r-1)(c-1)$ degree of freedom, r and c, respectively, are the number of rows and columns in the contingency table, or if the p-value is less than α, then reject H_0 at the α-level of significance; otherwise, fail to reject the H_0 (Walpole et al., 2012). According to the independent test obtained the Pearson's chi-squared statistics is $\chi^2 = 23.424$, with 6 degrees of freedom, for which the p-value is 0.001. Since the p-value is less than $\alpha = 0.05$, we have enough evidence to reject the null hypothesis.

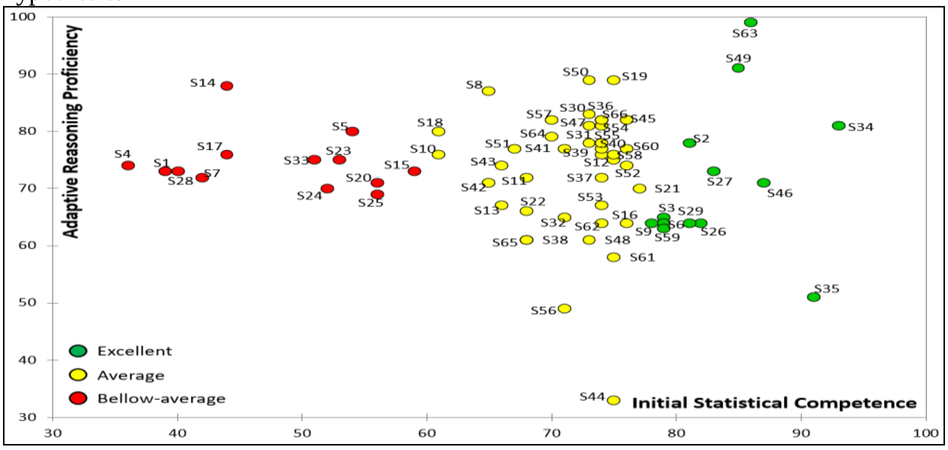


Figure 12

Scatter plot of the score on a preliminary test of Statistical Inference (x-axis) versus scattergram of the score on adaptive reasoning test (y-axis)

By correspondence analysis, the association is visualized graphically in the correspondence plot. Furthermore, the circle confidence regions for each point in a low-dimensional correspondence plot are also determined. These regions are used to identify whether a category is statistically consistent with what is expected under the H_0, based on the assumptions that underlie Pearson's chi-squared statistics (Beh & Lombardo, 2014; Lestari et al., 2019c).

The following two figures are correspondence plots that naturally depict the association structure among categorical variables. The symmetric plot is a plot of row and column principal coordinates in a joint display. The row-to-row and column-to-column distances are approximate chi-squared distances between the respective profiles. This plot is used to explore relationships among row (or column) categories and interpret the principal components related to the row (or column) categories. In this plot, the profiles are spread out to more easily view the distances between them. The row-to-column distances in the symmetric plot use two different mappings. Since the row-to-column distances are not standardized, thus the distances between row and column categories are not easy to interpret. Therefore, the asymmetric row plot is used to analyze these distances. Asymmetric plots allow us to intuitively explore the distances between row and column points, even if the two components represent a large proportion of the total inertia.

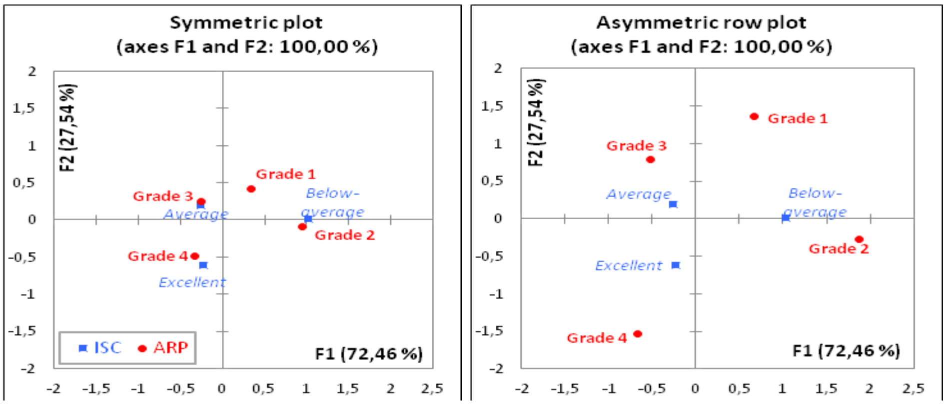


Figure 13
Symmetric correspondence plots (left) and asymmetric row correspondence plot (right)

To investigate which category contributes significantly to the association between variables can be known from the contribution biplot and elliptical confidence region as depicted in Figure 14. In contribution biplot, one set of points (usually row categorical variable) optimally represents spatial positions. The other set of points (usually column categorical variables) is represented by vectors related to their contributions to the low-dimensional plot (Greenacre, 2013).

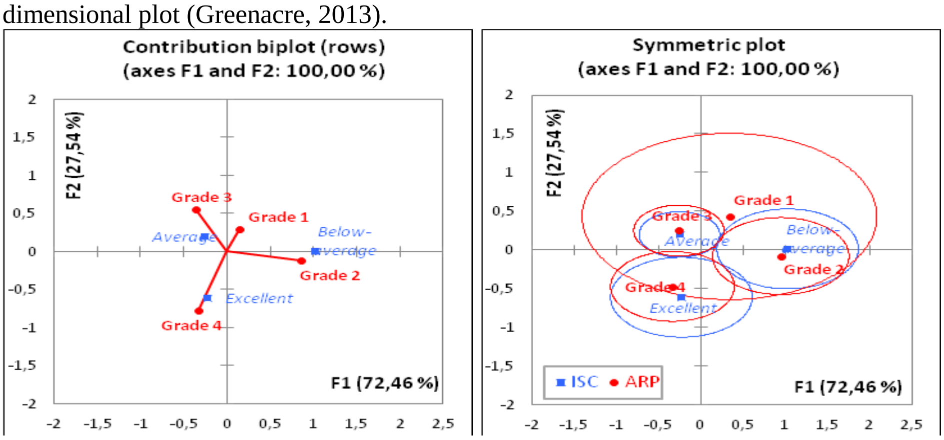


Figure 14
Contribution biplot (left) and elliptical confidence regions (right) in correspondence analysis (right)

Furthermore, to examine whether the correspondence plot represents the association, we can assess by constructing the elliptical confidence regions. Beh (2010) proposed these regions to investigate the statistical significance of categories that contribute to the overall association between categories. If the origin falls outside the ellipse, then the category is considered statistically important in defining the association structure between the categorical variables.

DISCUSSION

Classical statistical inference consists primarily of two procedures: hypothesis testing and confidence intervals. These techniques build on a scheme of interrelated concepts including probability, random sampling, parameter, distribution of values of a sample statistic, confidence interval, null and alternative hypothesis, p-value, significance level, and the logic of inference (Lui & Thompson, 2009). Consequently, adaptive reasoning is a fundamental reasoning process in statistical inference (Harradine et al., 2011).

Based on qualitative findings, the common problems in making statistical inference, most students may perform the calculations associated with an inferential process. Still, many students hold deep missing on the reasoning that prevents them from interpreting the result of an inferential process appropriately. This finding is in line with the results of previous studies (see Batanero, 2000; Castro-Sotos et al., 2007). It implies that students successfully implement hypothesis testing and parameter estimation procedures but lack explanation and justification of the result. Most of them fail to decide whether or not to perform a power analysis based on the previous testing. Previous research states that research on initial university statistics courses suggests that students have difficulty interpreting statistical inference appropriately (Makar & Rubin, 2009). Therefore, it required any prior experience in advanced algebra and calculus. The student also needed prior knowledge about sampling distributions to integrate several statistical concepts and to be able to reason about the hypothetical behaviour of many samples (Chance et al., 2004). This result indicates a link or association between initial statistical competence and adaptive reasoning. In order to explore this association, a quantitative study was conducted.

The quantitative findings imply a significant association between initial statistical competence and adaptive reasoning at the 95% confidence interval. Additionally, in

Figure 13, the first principal axis (x-axis) best explains row categories "below-average" and "excellent" and offers these two points farthest from the origin but with opposite signs. While the "average" is close to the origin, which indicates has a small contribution to the total variance (often called inertia) in the contingency table. The first axis contrasts the column category "grade 1" and "grade 2" with the column category

"grade 3" and "grade 4". The second axis (y-axis) contrasts the row categories "below-average" and "average" with "excellent", and the best explains the column category "grade 1". The asymmetric row plot in Figure 13 shows that row category "below-average" is relatively close to column category "grade 2", "average" close to "grade 3", and "excellent" close to "grade 4". This shows that there is a strong association among each pair of these categories. Overall, the first axis in the two plots above covers the largest variance is 72.46%.

The correspondence analysis solutions in Figure 14 show that the column category "grade 4" is the most important contribution in the association since it has the longest vector length. This grade also contributes strongly to the row categorical "excellent". On the other hand, "grade 1" gives only a few contributions since its vector length is shortest. The vector of "grade 1" is opposite to the vector of "grade 4", which indicates that "grade 1" and "grade 4" have a weak association. In addition, the vectors that represent "grade 1", "grade 2", "grade 3", and "grade 4" lie in different quadrants. It indicates that each grade's students' characteristics differ in their adaptive reasoning proficiency. Figure 14 (right) suggest that with the exception of "grade 3" and "average" (whose elliptical region overlaps the origin), all row and column categories contribute to the association structure between the two variables.

CONCLUSION

Qualitative findings describe that students' adaptive reasoning proficiency at each grade has different characteristics in meeting each indicator. The lowest grade meets these indicators less, while the highest grade successfully meets all indicators. The quantitative findings discover a significant association (dependence) between initial statistical competence and adaptive reasoning. The correspondence analysis solution depicts that a high level of statistical competence is strongly associated with a high grade of adaptive reasoning proficiency, and conversely. These results suggest that the mastery of initial statistical competence is an important aspect in developing students' adaptive reasoning proficiency. Therefore, to develop this proficiency in the Statistical Inference course, lecturers must pay attention to prior experience in advanced algebra or calculus and prior knowledge about sampling distributions.

ACKNOWLEDMENT

The authors would like to thank LPPM Universitas Singaperbangsa Karawang for supporting this research through grant research HIPKA UNSIKA 2021.

REFERENCES

Awofala, A. O. A. (2017). Assessing senior secondary school students mathematical proficiency as related to gender and performance in mathematics in Nigeria. *International Journal of Research in Education and Science*, *3*(2), 488-502. https://doi.org/10.21890/ijres.327908.

Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, *2*(1), 75-98.

Beh, E. J. (2010). Elliptical confidence regions for simple correspondence analysis. *Journal of Statistical Planning and Inference. 140*(9), 2582-2588. https://doi.org/10.1016/j.jspi.2010.03.018.

Beh, E. J., Lombardo, R., & Simonetti, B. (2010). A European perception of food using two methods of correspondence analysis. *Food Quality and Preference. 22*(2), 226-231. https://doi.org/10.1016/j.foodqual.2010.10.001.

Beh, E. J., & Lombardo, R. (2014). *Correspondence analysis theory, practice, and new strategies*. John Wiley & Sons, Ltd.

Castro-Sotos, A. E., Vanhoof, S., Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(1), 98–113.

Chance, B., Mas, R. D., & Garfield, J. (2004). *Reasoning about sampling distributions*. Springer.

Creswell, J. W. (2018). *Research design*. SAGE Publications Ltd.

Erickson, T. (2006). Using simulation to learn about inference. *Proceedings of the Seventh International Conference on Teaching Statistics*. Retrieved 22 October, 2021 from http://www.stat.auckland.ac.nz/~iase/publications/17/7G2_ERIC.pdf.

Gardner, H., & Hudson, I. (1999). University students' ability to apply statistical procedures. *Journal of Statistics Education*, *7*(1), 1-18.

Garfield, J. B. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*. *10*(3), 1-12. https://doi.org/10.1080/10691898.2002.11910676

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*. *2*(1), 22-38.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. Springer.

Greenacre, M. (2013). Contribution biplots. *Journal of Computational and Graphical Statistics. 22*(1), 107-122. https://doi.org/10.1080/10618600.2012.702494.

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*(5), 434-441.

Harradine, A., Batanero, C., & Rossman, A. (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education*. Springer.

Hembrough, T., & Jordan, J. (2020). Creating a digital writing classroom: A mixed methods study about a first-year composition tablet initiative. *International Journal of Instruction*, *13*(2), 567-586.https://doi.org/10.29333/iji.2020.13239a.

Hupe, J. M. (2015). Statistical inferences under the null hypothesis: Common mistakes and pitfalls in neuroimaging studies. *Frontiers in Neuroscience*. *9* (18), 1-9. https://doi.org/10.3389/fnins.2015.00018.

Kilpatrick, J., Swafford, J, & Findell, B. (2001). *Adding it up*. National Academy Press.

Lamnek, S. (2005). *Qualitative sozialforschung.* Psychologie Verlags Union.

Lepak, J. R., Wernet, J. L., & Ayieko, R. A. (2018). Capturing and characterizing students' strategic algebraic reasoning through cognitively demanding tasks with focus on representations. *The Journal of Mathematical Behavior*, *50*(6), 57-73.

Lesik, S. A. (2019). *Applied statistical inference with minitab*. CRC Press Taylor & Francis Group.

Lestari, K.E., Pasaribu, U.S., Indratno, S.W., & Garminia, H. (2019a). The comparative analysis of dependence for three-way contingency table using Burt matrix and tucker3 in correspondence analysis. *Journal of Physics: Conference Series*. *1245*(1), 012056. https://doi:10.1088/1742-6596/1245/1/012056.

Lestari, K.E., Pasaribu, U.S., & Indratno, S.W. (2019b). Graphical depiction of three-way association in contingency table using higher-order singular value decomposition Tucker3. *Journal of Physics: Conference Series*. *1280*(2), 022035. https:// doi:10.1088/1742-6596/1280/2/022035.

Lestari, K.E., Pasaribu, U.S., Indratno, S.W., & Garminia, H. (2019c). The reliability of crash car protection level based on the circle confidence region on the correspondence plot. *IOP Conference Series: Materials Science and Engineering*. *598*(1), 012061.

Lestari, K.E., Pasaribu, U.S., Indratno, S.W., & Garminia, H. (2020). Generating roots of cubic polynomials by cardano's approach on correspondence analysis. *Heliyon*, *6*(6), e03998. https://doi.org/10.1016/j.heliyon.2020.e03998.

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies: An International Journal*, *4*(2), 126-138.

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2021). Inferential statistical reasoning of math teacher: Experience in virtual contexts generated by the COVID-19 pandemic. *Education Science*. *11* (7), 363-382. https://doi.org/10.3390/educsci11070363.

Max, B., & Welder, R. M. (2020). Mathematics teacher educators' addressing the common core standards for mathematical practice in content courses for prospective elementary teachers: A focus on critiquing the reasoning of others. *The Mathematics Enthusiast*, *17*(2), 843-881.

McFeetors, P. J., & Palfy, K. (2018). Educative experiences in a games context: Supporting emerging reasoning in elementary school mathematics. *The Journal of Mathematical Behavior*, *50*(6), 103-125.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*. *8*(1), 82-105.

Nurjanah., Dahlan, J. A., & Wibisono, Y. (2021). The effect of hands-on and computer-based learning activities on conceptual understanding and mathematical reasoning. *International Journal of Instruction*, *14*(1), 143-160. https://doi.org/10.29333/iji.2021.1419a.

Ostler, E. (2011). Teaching adaptive and strategy reasoning through formula derivation: beyond formal semiotics. *International Journal of Mathematics Science Education*. *4*(2), 16-26.

Reichardt, C. S., & Gollob, H. F. (1997). When confidence intervals should be used instead of statistical significance tests, and vice versa. *In L. L. Harlow. S. A. Mulaik, & J. H. Steiger, What if there were no significance tests?*. Lawrence Erlbaum Associates.

Reid, N., & Cox, D. R. (2015). On some principles of statistical inference. *International Statistical Review*, *83*(2), 293-308.

Renchers, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis*. John Wiley & Sons Inc.

Risnawati, Andrian, D., Azmi, M. P., Amir, Z., & Nurdin, E. (2019). Development of a definition maps-based plane geometry module to improve the student teachers' mathematical reasoning ability. *International Journal of Instruction*, *12*(3), 541-560. https://doi.org/10.29333/iji.2019.12333a.

Schutt, R. (2014). *Investigating the social world: The process and practice of research*. Sage.

Staub, F. C., & Stern, E. (1997). Abstract reasoning with mathematical constructs. *International Journal of Educational Research*, *27*(1), 63–75. https://doi.org/10.1016/S0883-0355(97)88444-3.

Sotos, A.E.C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. Students' Misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Education Research Review*. *2*(2), 98–113. https://doi.org/10.1016/j.edurev.2007.04.001.

Víquez, H. A. (2022). Costa Rican preservice mathematics teachers' readiness to teach. *International Electronic Journal of Mathematics Education*, *17*(2), em0676.

Walpole, R. E., Myers, R. H., Myers, S. L., dan Ye, K. (2012). *Probability and statistics for engineers and scientist.* Pearson Education.

Yudhanegara, M. R., Indratno, S. W., & Sari, R. K. N. (2021). Dynamic item delivery network: prediction and clustering. *Heliyon*. *7*(5), e06934. https://doi.org/10.1016/j.heliyon.2021.e06934.

Yudhanegara, M. R., Indratno, S. W., & Sari, R. K. N. (2020a). Clustering for item distribution network. *Journal of Physics: Conference Series*.*1496*(1), 012019. https://doi:10.1088/1742-6596/1496/1/012019.

Yudhanegara, M. R., Indratno, S. W., & Sari, R. K. N. (2020b). Clustering for item delivery using rule-k-means. *Journal of The Indonesian Mathematics Society*. *26*(2),185-191. https://doi.org/10.22342/jims.26.2.871.185-191.

Yudhanegara, M. R., & Lestari, K. E. (2019). clustering for multi-dimensional data set: a case study on educational data. *Journal of Physics: Conference Series*. *1280*(4), 042025. https://doi:10.1088/1742-6596/1280/4/042025.