Jurnal Penelitian dan Pengukuran Psikologi
Volume 12, Nomor 02, Oktober 2023 http://doi.org/10.21009/JPPP.122.02

EVALUATION OF PSYCHOMETRIC CHARACTERISTICS OF A LOGICAL REASONING TOOL WITH THE CLASSIC TEST THEORY APPROACH

Dwi Yan Nugraha[1], Dian Nugraha[2], Widyastuti[3], Ahmad Ridfah[4]
Department of Psychology, Universitas Negeri Makassar[1,3,4]
Department of Mathematics Education Postgraduate Program, Universitas Negeri Makassar[2]

***E-mail**: dwiyannugraha20@gamil.com*

Abstract

Research on logical reasoning began to be carried out in Indonesia, it is necessary to conduct an evaluation of the psychometric characteristics possessed by logical reasoning measurement tools made by researchers. Logical reasoning measurement tools made by researchers need to be re-evaluated considering there are weaknesses in terms of using a relatively small number of subjects, so that it can affect the standardization of logical reasoning measuring devices. This study aims to obtain an overview of item difficulty index, item discrimination index, effectiveness of the distractors, validity, reliability, and error measurement from the logical reasoning measurement tool. Respondents in this study amounted to 7.730 students, consisting of 3.897 men and 3.833 women with an age range of 15-19 years. The research data were analyzed with the help of Microsoft Excel 2007, SPSS version 22.00, and ITEMAN version 3.6. The results of this study indicate that the item difficulty index moves from 0.35 to 0.80 which belongs to the easy and medium category. This study also shows that the item discrimination index moves from 0.10 to 0.71 with four items rejected and need to be aborted. This study also shows that the effectiveness of the distractors contained in each item in the logical reasoning measure is in the effective category. This study also shows that each item in the logical reasoning measure is considered valid in terms of factorial validity through exploratory factor analysis (EFA) procedures. This study also shows that the measurement of logical reasoning is relatively reliable with a Cronbach Alpha coefficient of 0.906. This study also shows that the measurement error obtained in this study is ± 3.92 of the total score obtained using a logical reasoning measure with a confidence level of 95%, with the actual score estimate obtained by respondents moving from 12.08 to 19.92, if the scores obtained by respondents sixteen.

Keywords: Classical theory, Evaluation, Logical reasoning measurement, Psychometric characteristics, Students

Abstrak

Penelitian tentang penalaran logis mulai banyak dilakukan di Indonesia, maka perlu untuk melakukan evaluasi mengenai karakteristik psikometrik yang dimiliki oleh alat ukur penalaran logis yang dibuat oleh peneliti. Alat ukur penalaran logis yang dibuat oleh peneliti perlu dievaluasi kembali mengingat terdapat kelemahan dari segi penggunaan jumlah subjek yang relatif sedikit, sehingga dapat mempengaruhi standarisasi pada alat ukur penalaran logis. Penelitian ini bertujuan untuk memperoleh gambaran indeks kesukaran item, daya diskriminasi item, efektivitas distraktor, validitas, reliabilitas, dan kesalahan pengukuran dari alat ukur penalaran logis. Responden dalam penelitian ini berjumlah 7.730 siswa, yang terdiri atas 3.897 laki-laki dan 3.833 perempuan dengan rentang usia 15-19 tahun. Data penelitian dianalisis dengan bantuan program Microsoft Excel 2007, SPSS versi 22.00, dan ITEMAN versi 3.6. Hasil penelitian ini menunjukkan bahwa indeks kesukaran item bergerak dari 0,35 sampai 0,80 yang tergolong pada kategori mudah dan sedang. Penelitian ini juga menunjukkan bahwa daya diskriminasi item bergerak dari 0,10 sampai 0,71 dengan empat item yang ditolak dan perlu digugurkan. Penelitian ini juga menunjukkan bahwa efektivitas distraktor yang terdapat pada setiap item dalam alat ukur penalaran logis berada pada kategori yang efektif. Penelitian ini juga menunjukkan bahwa setiap item dalam alat ukur penalaran logis tergolong valid ditinjau dari validitas faktorial melalui prosedur exploratory factor analysis (EFA). Penelitian ini juga menunjukkan bahwa alat ukur penalaran logis tergolong reliabel dengan koefisien Cronbach Alpha sebesar 0,906. Penelitian ini juga menunjukkan bahwa kesalahan pengukuran yang didapatkan dalam penelitian ini yaitu ±3,92 dari skor total yang diperoleh menggunakan alat ukur penalaran logis dengan tingkat kepercayaan 95%, dengan estimasi skor yang sebenarnya didapatkan oleh responden bergerak dari nilai 12,08 sampai 19,92, jika skor yang diperoleh oleh responden sebesar enam belas.

1. Introduction

Education in Indonesia was progressing and growing rapidly. The Indonesian state made education a means to improve the quality and ability of students. Pollat and Emre (2021) suggested that cognitive abilities and emotional skills were effective on student academic achievement. One of the cognitive abilities that had an impact on student's academic achievement was the ability to think logically. Balliel (2014) suggested that logical thinking was the willingness of students to solve problems with different operations. Students with logical thinking skills could see relationships between concepts as well as solve problems faced in everyday life. In addition, students can evaluate their thoughts, knowledge, and experiences by establishing cause-and-effect relationships (Ince Aka, 2012; Aydin, 2013).

In the field of education, there were also studies that examine the differences between logical thinking skills among students (Turgut et al., 2017; Hacıömeroğlu & Hacıömeroğlu, 2018; Yüzüak &; Dökme, 2019). For this reason, a scale was needed to measure students' logical thinking skills. Research on logical thinking measurement had been widely conducted, such as the group assessment of logical thinking [GALT] (Roadrangka et al., 1982), the 10-item logical thinking test (Tobin & Capie, 1981), which was used to present five reasoning situations (proportional reasoning, control variables, probability reasoning, relational reasoning, and combinational thinking). Then, the logical thinking ability scale from Lawson's classroom test of science reasoning (1978) consists of 16 multiple-choice questions. Furthermore, rational experiential inventory (REI) was developed by Pacini and Epstein (1999) and was applied to high school students (Türk & Artar, 2014). In addition, Nugraha (2017) had also constructed a logical reasoning ability instrument consisting of 29 multiple-choice questions.

The preparation of logical reasoning measuring instrument by the researcher was made in 2017 involving 100 respondents. The lack of respondents makes this a major weakness in standardizing measuring instruments. Based on these weaknesses, researchers needed to test psychometric characteristics on logical reasoning measuring instruments using a more significant number of subjects, so that future research related to logical reasoning can meet good standards. Coaley (2010) suggested that everything related to statistics and still using a small number of samples would impact data instability. The greater the number of samples used, the error in the SE mean would be decreased, and the observation mean would be closer to the actual mean of the population, and the results obtained would be more accurate. Coaley (2010) suggests that there are five criteria that are used as a basis for determining sample size, namely: 1) more than 2000 subjects classified as excellent, 2) 1000-1999 subjects classified as good, 3) 500-999 subjects classified as reasonable, 4) 200-499 subjects classified as adequate, and 5) under 200 subjects classified as inadequate. This study would seek to sample more than 2000 subjects so that it fell under the criteria of distinction. Therefore, due to the weaknesses in preparing this measuring instrument, it was considered necessary to evaluate the psychometric properties possessed by logical reasoning measuring instruments.

The evaluation of psychometric characteristics in this study uses a classical test theory approach (*CTT*). Classical test theory used a visible score measurement model, the sum of pure scores and measurement errors. Setiawati et al. (2018) suggest that the assumptions of the classical test theory approach are developed in various formulas that are useful in making psychological measurements. These formulas include difficulty level, differentiation power index, distraction effectiveness, test reliability, and measurement error. Based on the problems outlined in the background above, it was necessary to re-evaluate the psychometric characteristics of logical reasoning measuring instruments. The objectives of this study are to obtain: 1) an Overview of the item difficulty index from logical reasoning measuring instrument; 2) Overview of the discriminatory power of items from logical reasoning measuring instrument; 3) Overview of the effectiveness of the distractor of the logical reasoning measuring instrument; 4) An overview of the validity of logical reasoning measuring instrument; and 5) Overview of reliability and measurement error from logical reasoning measuring instrument.

2. Research Methods

Respondents

The number of respondents who had filled out the logical reasoning measuring instrument amounted to 7.730, consisting of 3.897 men and 3.833 women with an age range of 15-19 years (M = 16.98, SD = 1.42). Data that has been tabulated is then analyzed by performing data cleansing. Data was cleaned using boxplot facilities and extreme values in the SPSS program version 22.00. The facility was used to see outlier samples that affected the calculation results. Wardana (2007) suggested that outliers are extreme scores obtained by samples, both too low and too high, which impacted the results of data calculations. Samples that are above and below the horizontal line in the boxplot box indicate samples that are outliers and must be aborted. Sample scores at 5% of the highest and 5% of the lowest score on extreme values indicate extreme sample data, so the data must be invalidated. The data in this study was cleaned using the boxplot method contained in the SPSS program version 22.00.

Research design

This study used a quantitative approach in the field of psychometrics. Azwar (2015a) suggested that psychometrics was the science of procedures for evaluating the characteristics of psychological tests. This study evaluated the characteristics of logical reasoning measuring instruments made by the researcher. Evaluation of the characteristics in question, namely the level of difficulty of the item, the power of item discrimination, the effectiveness of distractions, validity, reliability, and measurement errors.

Instrument

Data was collected through a logical reasoning measuring instrument created by the researcher in 2017. This measuring tool comprised 29 questions with multiple answer options (A to E). To determine the level of logical reasoning ability of respondents, the total score was calculated by adding up the correct answers to each question. Higher scores indicated respondents who were good in aspects of logical reasoning. This logical reasoning measuring tool had an internal contingency value (α) of 0.81 (Nugraha, 2017). Although this measuring instrument already had an adequate reliability value, it still needed to be re-evaluated by involving a more significant number of respondents to obtain a more accurate standardization of measurement results.

Statistic analysis

The data in this study were analyzed using classical test theory approaches in evaluating the level of item difficulty, item discrimination power, distraction effectiveness, validity, reliability, and measurement error. Data analysis in this study used the help of Microsoft Excel 2007, SPSS version 22.00, and ITEMAN version 3.6.

3. Results

Preliminary analysis

The researcher first checked outlier data using the *boxplot* method, which reported that there were no outlier data, so as many as 7,730 respondents were used for further analysis, with 3,897 (50.41%) male students and 3,833 (49.58%) female students with an age range of 15-19 years, which was dominated by students aged 15 years (20.38%).

Psychometric properties

The researcher evaluated the psychometric properties of logical reasoning measuring instruments using parameters of item difficulty, item discrimination power, and distractor effectiveness with the help of ITEMAN application version 3.6. The results of the item difficulty index analysis reported that 16 items belonged to the easy category, 13 items belonged to the medium category, and no items belonged to the difficult category. Then, the results of the item discrimination power analysis showed that there were 25 items received and four items rejected, so the rejected items needed to be aborted. Furthermore, the results of the distractor effectiveness analysis showed that all answer choice values chosen by respondents were greater than the value of deceivers, so the options for answer choices contained in each item were in the effective category. The results of the evaluation of psychometric properties could be seen in the table (Table 1).

Factor analysis and reliability

The researcher also evaluated the validity and reliability of logical reasoning measuring instruments with factor analysis methods through exploratory factor analysis (EFA) procedures using the SPSS application version 22.00. The results of EFA analysis by rotating component matrix (Table 1) show that 17 items were included in factor 1, and 8 items were included in factor 2. The results of factorial validity analysis through the EFA procedure also showed the value of the factor charge on each factor formed. The factor load value was used to see the item's quality on the measuring instrument. The factor load value for each item formed in factors 1 and 2 is in the valid category because it was above the standard value (SLF > 0.30). In addition, the results of the reliability analysis (Table 1) also showed that the reliability value of Cronbach Alpha was 0.90 > 0.70, which means reliable. Analysis of measurement errors was performed using Microsoft Excel 2007 applications. Measurement errors were known as measurement errors. This measurement error could be obtained using the existing reliability value. Based on estimates with a 95% confidence standard, the value obtained is ±3.92 of the total score obtained using logical reasoning measuring tools. If the respondent got a score of 16, then the estimated score obtained by the respondent was from 12.08 to 19.92.

Table 1. Analysis Results Evaluation of Psychometric Properties

No	Item	*p*		*d*		Distractor		Factor (EFA)		Explanation
		Value	Category	Value	Category	Value	Category	SLF 1	SLF 2	
1	A1	0.75	Easy	0.10	Rejected	0.10	Effective	-	-	-
2	A2	0.35	Medium	0.56	Accepted	0.56	Effective	-	0.74	Valid
3	A3	0.46	Medium	0.67	Accepted	0.67	Effective	-	0.82	Valid
4	A4	0.58	Medium	0.71	Accepted	0.71	Effective	-	0.54	Valid
5	A5	0.80	Easy	0.21	Rejected	0.21	Effective	-	-	-
6	A6	0.61	Medium	0.70	Accepted	0.70	Effective	0.58	-	Valid
7	A7	0.59	Medium	0.71	Accepted	0.71	Effective	-	0.53	Valid
8	A8	0.48	Medium	0.68	Accepted	0.68	Effective	-	0.80	Valid
9	A9	0.41	Medium	0.62	Accepted	0.62	Effective	-	0.79	Valid
10	A10	0.46	Medium	0.58	Accepted	0.58	Effective	-	0.44	Valid
11	A11	0.49	Medium	0.69	Accepted	0.69	Effective	-	0.79	Valid
12	A12	0.65	Medium	0.71	Accepted	0.71	Effective	0.65	-	Valid
13	A13	0.72	Easy	0.50	Accepted	0.50	Effective	0.55	-	Valid
14	A14	0.68	Medium	0.56	Accepted	0.56	Effective	0.55	-	Valid
15	A15	0.74	Easy	0.46	Accepted	0.46	Effective	0.48	-	Valid
16	A16	0.74	Easy	0.47	Accepted	0.47	Effective	0.46	-	Valid
17	A17	0.74	Easy	0.43	Accepted	0.43	Effective	0.45	-	Valid
18	A18	0.73	Easy	0.43	Accepted	0.43	Effective	0.46	-	Valid

No	Item	*p*		*d*		Distractor		Factor (EFA)		Explanation
		Value	Category	Value	Category	Value	Category	SLF 1	SLF 2	
19	A19	0.73	Easy	0.47	Accepted	0.47	Effective	0.51	-	Valid
20	A20	0.74	Easy	0.46	Accepted	0.46	Effective	0.47	-	Valid
21	A21	0.73	Easy	0.47	Accepted	0.47	Effective	0.47	-	Valid
22	A22	0.73	Easy	0.46	Accepted	0.46	Effective	0.48	-	Valid
23	A23	0.74	Easy	0.45	Accepted	0.45	Effective	0.47	-	Valid
24	A24	0.73	Easy	0.47	Accepted	0.47	Effective	0.48	-	Valid
25	A25	0.73	Easy	0.46	Accepted	0.46	Effective	0.47	-	Valid
26	A26	0.73	Easy	0.44	Accepted	0.44	Effective	0.47	-	Valid
27	A27	0.74	Easy	0.48	Accepted	0.48	Effective	0.49	-	Valid
28	A28	0.49	Medium	0.20	Rejected	0.20	Effective	-	-	-
29	A29	0.47	Medium	0.20	Rejected	0.20	Effective	-	-	-
Cronbach Alpha (α) = 0.90 > 0.70										Reliable

Note: *p*= Item difficulty index; *d* = Item discrimination power; SLF = Standardized loading factor (SLF > 0.30).

4. Discussion

This study obtained the difficulty index value of items moved from 0.35 to 0.80. Azwar (2016) suggested that the item difficulty index was a parameter that describes the difficulty of an item for a group of subjects given a test to give the correct answer. The difficulty index of the item was symbolized by the letter *p*. A high p-value indicated that the item was easier to answer, while a low p-value indicated that the item was getting harder for the group of subjects concerned. Gregory (2013) suggested that the criteria for item difficulty index values were divided into three categories, namely 1) $p < 0.30$ was in the difficult category. 2) $0.30 < p < 0.70$ was in the medium category. 3) $p > 0.70$ was in the easy category. The results of the item difficulty index obtained in this study show that the items contained in the logical reasoning measuring instrument are in the easy and medium categories.

This study showed that the value of the discrimination power of the obtained items moved from 0.10 to 0.71. Azwar (2016) suggested that item discrimination power is the extent to which an item's ability to distinguish individuals from one another based on attributes measured by tests. The discriminatory power of items was reflected by the difference in answers to each item between intelligent and unintelligent groups of subjects. An intelligent subject group had a greater chance of correctly answering an item compared to another group of subjects. Azwar (2012) suggested that the criteria for evaluating item discrimination power used a minimum coefficient limitation of 0.30. The value of the item discrimination power produced on this measuring instrument shows that four items did not meet the criteria, namely 1, 5, 28, and 29, so these items are rejected and need to be aborted.

Azwar (2016) suggested that good items had high item discrimination, appropriate difficulty, and effective distractors. Hartono (2015) suggested that the effectiveness of the distractor was said to be effective if the value of the answer choice chosen by the respondent was greater than the value of the deceiver. Azwar (2015b) suggested that the effectiveness of the distractor was used to see that the distractor functioned as it should; that was, the distractor was chosen by more low groups while from the high group, only a few or no choices. The effectiveness of the distractor in this study could be seen from the value of the Alternative Statistics section of Point Biser, which showed that all the value of answer choices chosen by respondents was greater than the value of the deceiver so that the options for the answer choices contained in each item are in the effective category. Furthermore, this study used the EFA method as part of the construct validity regarding the highest factor charge value. Hair et al. (2014) suggested that a factor load value above 0.30 indicated that the item could better explain the construct and was declared valid and significant. This study showed that the items on the logical reasoning measuring instrument were divided into two factors, and each item on the measuring instrument is in the valid category because it is above the value of 0.30.

The reliability results in this study showed that logical reasoning measuring instruments were classified as reliable. The Cronbach Alpha coefficient obtained was 0.90. Kaplan and Saccuzzo (2012) suggested that the value of the reliability coefficient above 0.70 indicated that the measuring instrument was reliable. The measurement results of a measuring instrument could not be separated from the measurement error. Setiawati et al. (2018) suggested that reliability and measurement error were two interrelated things. Suryabrata (2005) suggested that the higher the reliability coefficient of a test, the smaller the possibility of measurement errors. This measurement error could be obtained using the reliability values generated on logical reasoning measuring instruments. Based on estimates with a 95% confidence standard, the value obtained is ±3.92 of the total score obtained using logical reasoning measuring tools. If the respondent got a score of 16, then the estimated actual score obtained by the respondent was from 12.08 to 19.92. These results also provide additional evidence regarding the meaning of reliability coefficients in logical reasoning measuring instruments.

5. Conclusion

Based on the results of research on the evaluation of psychometric characteristics of logical reasoning measuring instruments with a classical test theory approach, it could be concluded that the difficulty index of items obtained in this study moves from 0.35 to 0.80, so that the items contained in the logical reasoning measuring instrument are in the easy category as many as 16 items, and the medium category as many as 13 items, and none of the items fall into the difficult category. Then, the value of the descriptive power of the items obtained in this study moved from 0.10 to 0.71, and four items did not meet the criteria, namely items 1, 5, 28, and 29, so the items were rejected and needed to be aborted. Furthermore, all answer choice values chosen by respondents were greater in percentage than the value of deceivers, so the options for answer choices contained in each item were in the effective category. In addition, there are two factors (factor 1 = 17 items and factor 2 = 8 items) formed on logical reasoning measuring instrument with factor load values for each item formed in factors 1 and 2 were in the valid and reliable category, with a Cronbach Alpha reliability value of 0.90. The study also found a measurement error rate in this study of ±3.92 of the total score obtained using logical reasoning measuring tools, with an estimated confidence of 95%. If the respondent got a score of 16, then the estimated actual score obtained by the respondent was from 12.08 to 19.92. Suggestions for future research were expected that researchers improve the quality of items that had low item discrimination power and name the two factors that make up logical reasoning measuring instruments, develop norms on logical reasoning measuring instruments, and analyze psychometric characteristics again using modern theoretical approaches in the Rasch model.

6. References

Aydın, Ö. (2013). *Effectiveness of argumentation (discussion theory) in science and technology teacher candidate education*. Dissertation. Hacettepe University Graduate School of Social Sciences.

Azwar, S. (2012). *Reliabilitas dan validitas*. Pustaka Pelajar.

Azwar, S. (2015a). *Dasar-dasar psikometrika (Edisi 2 cetakan I)*. Pustaka Pelajar.

Azwar, S. (2015b). *Tes prestasi: Fungsi dan pengembangan pengukuran prestasi belajar*. Pustaka Pelajar.

Azwar, S. (2016). *Konstruksi tes kemampuan kognitif (Edisi 1)*. Pustaka Pelajar.

Ballıel, B. (2014). *The effect of webquest supported cooperative learning approach to learning products. Dissertation.* Gazi University Graduate School of Educational Sciences.

Coaley, K. (2010). *An introduction to psychological assessment and psychometrics*. SAGE Publications.

Gregory, R. J. (2013). *Tes psikologi: Sejarah, prinsip, dan aplikasinya*. Erlangga.

Hacıömeroğlu, E. S., & Hacıömeroğlu, G. (2018). Examining prospective teachers' logical reasoning ability: The Longeot's test of cognitive development. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 9*(3), 413–448. https://doi.org/10.17762/turcomat.v9i3.183

Hair J. F., Black, W. C., Babin, B. J., & Anderson R. E. (2014). *Multivariate data analysis* (7th. Ed.). Pearson Education Limited.

Hartono. (2015). *Analisis item instrumen (Cetakan I)*. Zanafa Publishing.

İnce Aka, E. (2012). *The effect of problem-based learning method used for teaching acids and bases on different variables and students' views on the method*. Dissertation. Gazi University Graduate School of Educational Sciences.

Kaplan, R. M., & Saccuzzo, D. P. (2012). *Pengukuran psikologi: Prinsip, penerapan, dan isu (Edisi 7)*. Terjemahan oleh E. P. Widodo. Salemba Humanika.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15, 11–24. https://doi.org/10.1002/tea.3660150103

Nugraha, D. Y. (2017). *Laporan konstruksi alat ukur psikologi (Alat ukur penalaran logis)*. Tugas Akhir Mata Kuliah Konstruksi Alat Ukur Psikologi tidak diterbitkan. Program Studi Psikologi Fakultas Psikologi Universitas Negeri Makassar.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972– 987. https://doi.org/10.1037/0022-3514.76.6.972

Polat, H., & Emre, F. B. (2021). Logical Thinking Abilities Scale: Reliability and Validity Studies. *Universal Journal of Educational Research*, *9*(1), 180–185. https://doi.org/10.13189/ujer.2021.090120

Roadrangka, V., Yeany, R., & Padilla, M. (1982). *GALT, Group test of logical thinking*. University of Georgia.

Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018). Evaluasi karakteristik psikometrik tes bakat differensial dengan teori klasik. *Jurnal Humanitas*, *15*(1), 46–61. https:// doi.org/10.26555/humanitas.v15i1.7249

Suryabrata, S. (2005). *Pengembangan alat ukur psikologis (Edisi 3)*. ANDI.

Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological measurement*, *41*(2), 413–423. https://doi.org/10.1177/001316448104100220

Turgut, M., Yenilmez, K., & Balbağ, M. Z. (2017). Öğretmen adaylarının mantıksal ve uzamsal düşünme becerileri: bölüm, cinsiyet ve akademik performansın etkisi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, *1*(41), 265–283. https://doi.org/10.21764/efd.13098

Türk, G. T., & Artar, M. (2014). Adaptation of the rational experiential inventory: Study of reliability and validity. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, *47*(1), 1–18. https://doi.org/10.1501/Egifak_0000001314

Wardana, A. (2007). *Modul metode penelitian sosial budaya (Menggunakan SPSS dalam penelitian sosial)*. Program Studi Pendidikan Sosiologi Jurusan Pendidikan Sejarah Fakultas Ilmu Sosial dan Ekonomi Universitas Negeri Yogyakarta.

Yüzüak, A. V., & Dökme, İ. (2019). Science reasoning levels of prospective science and primary teacher. *Hacettepe University Journal of Education*,*34*(3), 586–601. https://doi.org/10.16986/huje.2018046314