

Brief Report

Argument Evaluation Test for Critical Thinking Ability Assessment: A Preliminary Study

Christin N. K. Lunggito, Anindito Aditomo, and Ide Bagus Siaputra
Faculty of Psychology
Universitas Surabaya

Critical thinking is a highly valued learning outcome in all educational levels. This study explores students ability to evaluate arguments as one manifestation of critical thinking. Using a newly developed instrument (the Argument Evaluation Test), we describe high school graduates' ($N = 2201$) ability to distinguish between sound and fallacious arguments, without explicit instruction. The results indicate that about half of the participants were able to spontaneously evaluate arguments and identify logical fallacies. This is significant considering that argumentation and logical fallacies are not part of the school curriculum in Indonesia. Thus, the ability to identify at least some types of logical fallacies could be obtained through informal learning and therefore the knowledge underlying this ability is likely to be implicit. Another finding was that the ability to evaluate arguments seem to vary depending on the type of fallacy, with ad hominem arguments easier to be identified as fallacious compared to ad populum arguments, as well as demographic variables such as gender and ethnicity

Keywords: critical thinking, argument evaluation test, fallacious reasoning

Keterampilan berpikir kritis adalah capaian pembelajaran yang sangat penting. Penelitian ini mengeksplorasi salah satu bentuk spesifik keterampilan berpikir kritis, yakni kemampuan mengevaluasi argumen. Menggunakan sebuah instrumen baru (Tes Evaluasi Argumen), peneliti memetakan tingkat kemampuan evaluasi argumen remaja lulusan SMA ($N = 2201$) yang akan menjadi mahasiswa di sebuah perguruan tinggi swasta. Hasil penelitian menunjukkan bahwa cukup banyak lulusan SMA yang secara spontan mampu mengevaluasi argumen. Temuan ini cukup signifikan, mengingat argumentasi dan kesalahan penalaran (*logical fallacies*) tidak diajarkan secara formal di sekolah dan universitas di Indonesia. Dengan demikian, kemampuan mengevaluasi argumen dan mengenali kesalahan penalaran tampaknya didapat melalui proses pembelajaran yang informal dan karenanya pengetahuan tersebut bersifat implisit. Simpulan lain yang bisa diperoleh adalah bahwa kemampuan mengevaluasi argumen tampaknya bervariasi, tergantung pada jenis kesalahan penalaran yang harus dievaluasi serta variabel demografis seperti gender dan etnis.

Kata kunci: berpikir kritis, tes evaluasi argumen, kesalahan penalaran

Reasoning is a specific learning achievement of every course of study, apart from mastery of knowledge and work skills. Reasoning is also a general ability required in any relevant situations. Since reasoning is a general ability, any formal education put this ability as main objectives, particularly in higher education (Kuhn, 2005). The importance of reasoning ability agreed by those who

work with university graduates (Badcock et al, 2010). Google, for instance, had decided that the first requirement for their new recruited employees was new information learning process skill for problem solving. Google did not make the GPA nor any other academic achievement at school as the employee recruitment main criterions (Friedman, 2014).

The faculties and the higher education authority realized the importance of general thinking ability. This was clearly seen from their assertion about desired university graduates qualification. The government of

This research was funded by a Fundamental Research grant (2014-2015) from the Indonesian Directorate of Higher Education to the second and third authors.

Correspondence concerning this article should be addressed to Anindito Aditomo, Faculty of Psychology, Universitas Surabaya, Jalan Raya Kalirungkut, Surabaya 60293. E-mail: a.aditomo@gmail.com

Indonesia, through its Culture and Education Ministry, defined that learning aims for higher education programs were not only the mastery of knowledge and several specific skills but also general ability such as logical, critical thinking and systematic thinking (Permendikbud 49/2014).

This awareness of the importance of general thinking ability was supportive but it was also critical to note that thinking ability did not automatically improve through the lectures delivered in classes. The research held by Perkins (1985) for instance, showed that individual's level of education did not influence the skill of argument analysis. Recent research held by Arum and Roksa (2011) showed similar results. Unfortunately, researches on general thinking development among students in Indonesia were still a few. Evidences showed that the learning process in schools in Indonesia was still trying hard to develop students' reasoning ability (Musyahid, 2009).

This writing was going to report a research finding on general thinking ability among students in Indonesia universities. This finding would give a systematic understanding about this specific ability. This preliminary research focused on analytical thinking and arguments evaluation ability. This two kinds of specific abilities were part of critical like thinking skill. Critical thinking was defined as a purposeful thinking process which was monitored and supervised to solve problems, to make decisions or to learn new concepts (Evens, Verburch & Elen, 2013). In other words, critical thinking involved meta-cognitive thinking process. This process was an awareness of mental process quality and its products (Moshman, 2011). Argumentation was an application of critical thinking process to analyze claims or arguments (Brem, Russel & Weems, 2001). Argumentation or arguments analyzes was defined as an assessment of strengths and weaknesses of an assertion or argument of current issues (Perkins, 1985).

During an argument analyzing process, one needs to be alert of possibilities that the strong-looking assertions had weaknesses. These kind of deceiving arguments often appeared like *informal reasoning fallacy* (Neuman et al, 2006). An argument would contain reasoning fallacy if it violated certain rationalities (Walton, 2010, p.160). For example, a cosmetic advertisement claimed that "nine out of ten people" chose the product. The ad said implicitly that the cosmetic product was qualified and worth buying. The ad seemed to contain a piece of beneficial information for consumers, gathered from survey data. The survey was probably accurate, but the product popularity could not be used to support the quality claim. This is an example of *ad populum fallacy*,

a reasoning fallacy in which an argument made the most of common agreement or public opinion to convince someone about something (Walton, 2008).

Another kind of informal reasoning fallacy was *ad hominem* arguments. This arguments were the one used to criticize the messenger and did not denounce the claim nor the arguments' contents itself (Mizrahi, 2010). In other words, *ad hominem* arguments was employed to criticize the messenger's character and ignore the contents of the arguments. *Ad hominem* arguments were classified as reasoning fallacy. Although the messenger had bad characters, this did not automatically weakening the arguments.

Based on the concept of critical thinking, a critical thinker should be able to consciously analyzed the thinking process and employed rational standard to identify a fallacious argument. Based on this consideration, one would presumably had argumentation ability if she/he had the skill to distinguish the strong-and-logical arguments and the fallacy-contained arguments. As far as the researcher's knowledge, there were still few researches on argumentation ability in Indonesia. This research was held to examine analyzing arguments ability among students of higher education in Indonesia.

The related research on argument analysis was once held by Stanovich dan West (1997). In order to measure argument analysis ability, they used a test which provided diverse qualification arguments on controversial topics. Every single arguments was then scored and the score represented the argument quality (according to expert judgments). Through this research Stanovich dan West developed measurement technique. They included various kinds of arguments contained certain mistakes, such as *argumentum ad populum* and *ad hominem*. They included *argumentum ad populum* and *ad hominem* since these two kind of mistakes were commonly found and frequently effective to mislead one's reasoning judgments (Walton, 2008).

Based on the above explanation, this research was going to find out the critical thinking ability among a group of high school graduates who were about to be admitted in a private university. Specifically, this research was carried to describe the number of high school graduates who were able to recognize *ad hominem* and *ad populum* arguments implicitly presented to them (in other words, without any specific instruction to find the logical fallacies and without any definition provided for each arguments). This method was important since critical thinking ability should be expressed spontaneously in a situation in order to be measured. Besides, this research was also held to find if there was any possibility of argument evaluation

skills variance across demographic variables (gender, ethnicities, parents' background).

Method

Research Design

This research was a cross sectional quantitative survey. The participants were high school graduates who enrolled to be admitted as students of a private university in Surabaya – Indonesia. The data were gathered through questionnaires distributed during the university orientation program for new students. Following this program was a mandatory for every new student in the university.

Participants

There were 2201 new students participated in this research. They were about 88 % of total new admitted students population. Their mean age was 17.96 years old and mostly (75%) were females. Most of them identified themselves as Chinese (45.7%), Javanese (36.8%), and about 14.5 % of them were of Bugis, Batak, Bali, Dayak, Sunda, Madura and other kind of ethnicities. Their parents' educational background was presented in Table 1.

Variables and Measurement Instrument

Demographic data were obtained from the open questions presented at the few first pages of the questionnaire. The arguments evaluation ability was the main variable of this research, which was defined as the ability to recognize arguments with reasoning fallacy. This variable was measured by Argument Evaluation Test developed by the second author (Anindito Aditomo), based on the adapted instrument developed by Stanovich and West (1997). The Argument Evaluation Test con-

tained of controversial social/political policies such as the policy of smoking prohibition in public places and women movement restriction in a region. For each policies there were two sets of dialogue started with a short argument. This argument criticized the policies and then followed by three rebuttal toward the argument. Two of these three rebuttals contained *ad hominem* dan *ad populum* reasoning fallacies, and the other one was a substantive argument (referring to the factual issue criticized in the earlier argument).

The Argument Evaluation Test was developed to measure the spontaneous critical thinking ability of an individual in daily situations, without any instructions nor explicit directions to think critically. Therefore, in administering the Argument Evaluations Test the participants were not explicitly asked to find which arguments was substantial and which contained reasoning fallacy. But they were asked to rate the strength and weaknesses (based on scale 1 to 6) of each arguments. The score of argument evaluation ability gained from the difference between the score of substantive argument and the score of fallacy-contained argument. Those with spontaneous critical thinking ability would give lower score to both *ad hominem* argument and *ad populum* argument than to substantive argument.

For the purpose of this research, the Argument Evaluation Test accomodated three cases/ issues and each case equipped with two sets of argumentative dialogues; *ad hominem* and *ad populum* arguments. Therefore, each *ad hominem* and *ad populum* sub tests had six items. One section of the Argument Evaluation Test was included in the attachment, and the complete test version could be obtained from the second author (Anindito Aditomo).

Analysis

Right after tabulation and data cleaning, items realibility test of the Arguments Evaluation Test was completed. The reliability test for items of *ad hominem* arguments was held apart from the test for items of *ad populum* arguments. The reliability test results indicates that each sub-test had adequate internal consistency (Alpha Cronbach .69 for items of *ad hominem* sub-test dan .68 for items of *ad populum* sub-tes). The coefficient of corrected item-total correlation for each sub-test was more than 0.3. And this meant that there was not any items that was too diverse from the others (Azwar, 1996). When the reliability analysis was done, the descriptive analysis was held to get mean score and classification of argument evaluation ability. The Kolmogorov-Smirnov test showed that the data distri-

Table 1

Parents' Educational Level

Level of Education	Father		Mother	
	N	%	N	%
Elementary to Middle Level of Education	1212	55.1	1360	61.8
Undergraduate Level of Education	803	36.5	762	34.6
Master Level of Education	184	8.4	76	3.5

bution for the two sub-tests was not normally allocated. Consequently, the Mann-Whitney test and Kruskal-Wallis test was used to find the difference of argument evaluation skill scores among demographic groups.

Results and Discussion

The argument evaluation score obtained from the difference of substantive-arguments rebuttal appraisals scores and each fallacious argument score; *ad hominem* reasoning fallacy score and *ad populum* reasoning fallacy score. The rebuttals score range is 1 (weak) to 6 (strong), consequently the argument evaluation ability score, for each sub-scale was between -5 to 5. The positive scores represented the ability to recognize fallacious arguments as weaker arguments than substantive ones. The null and negative scores depicted the participants' failure to recognize reasoning fallacy arguments.

The scores showed that the research participants' argument evaluation abilities for *ad hominem* sub-scale was 1.44 ($SD = 1.26$), the average score for *ad populum* sub-scale was 0.66 ($SD = 1.26$). This depicted that the participants generally were able to distinguish substantive arguments and *ad hominem* and *ad populum* arguments. Besides, they were more skillful to evaluate *ad hominem* fallacious arguments than *ad populum* ones. In other words, offensive arguments were considered weak arguments. This was interesting since items of *ad hominem* arguments in the Argument Evaluation Test were not offensive nor impolite. Therefore, the low scores given to the items of *ad hominem* arguments was not based on politeness norms.

In order to get better description of the participants' argument evaluation ability, their scores were classified. The classification criteria were based on the decision whether the participants were able to distinguish substantive arguments and reasoning fallacious ones, regardless of 'the differences distance.' For example, person A scored 3 (rather weak) to a substantive argument and scored 2 (weak) to a fallacious argument, consequently, she/he would be in the same classification as person B, who gave 6 (very strong) to a substantive argument and 1 (very weak) to a fallacious argument. Since there were two sub-tests (*ad hominem* and *ad populum*), three classifications of argument evaluation ability were created and each was defined as follows: (1) High; when the participants consistently recognize *ad hominem* and *ad populum* arguments (mean score for the two sub-scales was bigger than null); (2) Medium: when the participants recognize one of the *ad hominem* and *ad populum* arguments (mean score

for one of the two sub-scales was bigger than null); (3) Low; when the participants did not recognize two kinds of reasoning fallacious argument (mean score for the two sub-scales was null or negative).

When the scores obtained were classified, it depicted that many participants (49.7%) were highly able to evaluate arguments. About half of the high school graduates were spontaneously able to recognize and distinguish *ad hominem*, *ad populum* and substantive arguments. There was 16.2 % of the participants showed low level of argument evaluation ability. They did not recognize any *ad hominem* and *ad populum fallacies* in an argument and they showed difficulties distinguishing the fallacies with substantive arguments. About one third (34.1%) of them was only able to recognize one of the two kinds of reasoning fallacies.

The researchers also tried to find if there was any differences among gender, ethnicities and parents' educational background in argument evaluation ability. The data analysis indicated that female participants were more skillful recognizing *ad hominem* and *ad populum* arguments than men. Minority ethnic groups ("others" classification) were more able to recognize *ad hominem* arguments than Javanese and Chinese. Parents' educational level was not an influential variable for argument evaluation ability.

Limitations

Some research limitations issues presented here were to be considered in order to value the research findings. First, the Argument Evaluation Test applied in this research contained only two kinds of reasoning fallacy. Considering that the difficulty level of the two fallacies was different, the various kinds of arguments and reasoning fallacies should be accommodated in the next version of the test. Secondly, it was not known yet if responses to Reasoning Evaluation Test were predictive of learning outcomes such as GPA or other theoretically related psychological variables such as intelligence. Third, the data was gathered from the participants with relative same age and educational background. The next similar research need to involve participants with various age, level of education and work experiences.

Conclusions and Suggestions

The research found that many high school graduates, who were about to study in universities, were spontaneously able to evaluate arguments. This finding was significant enough as arguments analysis and *logical*

fallacies were not part of formal learning subjects in Indonesia's schools and universities. Therefore, the ability of argument evaluation and reasoning fallacies recognition were likely acquired from informal learning process dan learned implicitly. Another finding was the various level of argument evaluation ability which was affected by the kinds of reasoning fallacies to be evaluated and demographic variables such as gender and ethnicities. These findings would need further research in order to get better understanding.

References

- Arum, R., & Roksa, J. (2010). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.
- Badcock, P. B. T., Pattison, P. E., & Harris, Kerrie-Lee. (2010). Developing generic skill through university study: a study of arts, science, and engineering in australia. *High Edu*, 60, 441-458. doi: 10.1007/s10734-010-9308-8
- Brem, S. K., Russel, J., & Weems, L. (2001). Science on the web: student evaluations of scientific arguments. *Discourses Processes*, 32(2&3), 191-213.
- Evens, M., Verburch, A., & Elen, J. (2013). Critical thinking in college freshmen: The impact of secondary and higher education. *International Journal of Higher Education*. 2(3), 139-151. doi: 10.5430/ijhe.v2n3p139
- Friedman, T. L. (2014). *How to get a job at google*. Retrieved from http://www.nytimes.com/2014/02/23/opinion/sunday/friedman-how-to-get-a-job-at-google.html?_r=0
- Kuhn, D., & Park, S.H. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, 4, 111-124.
- Mizrahi, M. (2010). Take my advice-I am not following it: *Ad hominem* arguments as legitimate rebuttals to appeals to authority. *Informal Logic*, 30(4), 435-456.
- Moshman, D. (2011). *Adolescent, rationality and development: Cognition, morality, and identity* (3rd ed.). New York: Taylor and Francis Group.
- Neuman, Y., Weinstock, M.P., & Glasner, A. (2006). The effect of contextual factors on the judgement of informal reasoning fallacies. *Quarterly Journal of Experimental Psychology*, 59(2), 411-425.
- Permendikbud Nomor 49 (2014). *Standar Nasional Pendidikan Tinggi*. Retrieved from <http://www.kopertis12.or.id/2014/06/11/pemendikbud-no-49-tahun-2014-tentang-standar-nasional-pendidikan-tinggi.html>
- Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, 77(5), 562-571.
- Pujiastuti, P. (n.d). *Pembelajaran kreatif-produktif untuk mengembangkan kemampuan berpikir kritis dan kreatif bagi mahasiswa*. Retrieved from [http://staff.uny.ac.id/sites/default/files/penelitian/Dr.%20Pratiwi%20Puji%20Astuti,%20%20M.Pd./PEMBELAJARAN%20KREATIF-PRODUKTIF%20\(%20Artikel%20Pratiwi\).pdf](http://staff.uny.ac.id/sites/default/files/penelitian/Dr.%20Pratiwi%20Puji%20Astuti,%20%20M.Pd./PEMBELAJARAN%20KREATIF-PRODUKTIF%20(%20Artikel%20Pratiwi).pdf)
- Stanovich, K. E., & West, R. F. (1997). Reasoning independently prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, 89 (2), 342-357
- Walton, D. (2008). *Informal logic: A pragmatic approach* (2nd Ed). New York : Cambridge University Press.
- Walton, D. (2010). Why fallacies appear to be better arguments than they are. *Informal Logic*, 30(2), 159-184.