International Journal of Learning, Teaching and Educational Research
Vol. 19, No. 11, pp. 223-241, November 2020
https://doi.org/10.26803/ijlter.19.11.13

Shadow Education in Indonesia: Is It Relevant to Students' Critical Thinking Skills in Chemistry Learning?

Maria C. S. Kawedhar, Sri Mulyani, Sulistyo Saputro and Sri Yamtinah
Universitas Sebelas Maret, Surakarta, Indonesia
https://orcid.org/0000-0002-1669-8699
https://orcid.org/0000-0002-0181-1345
https://orcid.org/0000-0002-5613-9496
https://orcid.org/0000-0001-5616-4897

Abstract. Private tutoring, also known as shadow education (SE), is significant in the view of Indonesian society. It attains public reliance on children's success at school. Although SE seems to be promising to elevate students' achievements in school, this still needs to be proven. With the development of 21st-century skills, can SE make itself worthy of being a part of proper supplementary education? This research aims to reveal how relevant SE is to develop students' 21st-century skills, especially critical thinking (CT) skills. This study involved 193 students from 3 different ranks of schools in Surakarta, Indonesia. This mixed methods explanatory research used a computerized testlet instrument to obtain quantitative data about students' CT skills in Chemistry. Open-ended questionnaires and open-ended interviews were used to obtain qualitative data about students' tendencies in their SE. Interviews were also performed to hear SE tutors' voices about the SE field in which they worked. The quantitative data were analyzed using Rasch modeling to reveal the interaction between the items and students' CT skills. The modeling results showed that SE still seems less significant in responding to the community's wishes about their children's success at school, especially as related to CT skills. However, the results also showed that SE can aid those with moderate cognitive ability. Furthermore, the qualitative results are worth considering by those who will take or build SE in future. Thus, SE can be a proper supplement for formal education.

Keywords: shadow education; critical thinking; Rasch

1. Introduction

Indonesia, a country in Southeast Asia, has a significant population, occupying the 4th rank globally. The United Nations (UN) estimates that Indonesia reached 274,827,175 citizens in early December of 2020 (Worldometer, 2020). The data

provided by the Data Center of Education and Culture Ministry of Indonesia show that by late December 2020, the total number of students registered and still active was 52,807,275 (Directorate General of Early Childhood, Primary, and Secondary Education, 2020). This situation requires strict attention by educators and the government. The 2018 PISA database showed that Indonesia stood in 71st place out of 76 countries in reading comprehension and 69th out of 77 countries in Science performance (OECD, 2018). This significant number of both students and problems present an excellent opportunity for Indonesia's education to develop and be fixed. Progress may occur in terms of the number of educational material and facility providers.

Competition among students is also becoming more challenging. Large numbers of students mean that students have to scramble for placement in bona fide schools. Although Indonesia is one of the countries with the highest number of private and public schools, private schools still cannot lower the level of competition among students and public schools remain the destination for many students. This intense competition has resulted in students and parents making various attempts and putting in more effort to get curated at their favorite school. One of the efforts is for students to attend a private tutoring course that takes on the role of supplementary education and which is commonly referred to as shadow education (SE) (Bray, 2013). SE is great for parents worrying about their children's ability to compete at school or to obtain higher scores. Therefore, many parents trust SE as a helper and a complement to formal education (Bray, 2014). SE follows the development of education around the world and is spreading worldwide (Nam & Chan, 2019). However, it does not focus on simple cognitive aspects. Nowadays, students need more complex abilities to succeed in their cognitive abilities. Skills are what they need as the key to working readiness (Krishnakumar & Nogales, 2020). Together with knowledge and abilities, skills have equal importance for students to complete their course (Mulyani, Widiastuti & Supriyanto, 2019).

1.1 Literature Review

Critical thinking (CT) skills, as the name suggests, become more critical in this accelerating world of change. Students are forced to face new problems, new learning styles, new adaptation, new opportunities, new ways of thinking, and even new recipes to produce more effective and efficient learning solutions. Obtaining CT skills is not as easy as having a sweet dream, though. It requires a long and integrated process (Dwyer, Hogan & Stewart, 2014). CT is very important, especially in Chemistry learning, since it plays a big role in students' ability to gain a complex understanding of learning. There is enough evidence showing that Chemistry is included in the "tough discipline" group. Many students in some countries have tried to study Chemistry but still failed (Woldeamanuel, Atagana & Engida, 2014). This indicates that CT skills are highly needed in Chemistry learning to address this problem. Furthermore, Mahaffy (2004) has declared the prospective form of Chemistry, making Chemistry more complex since it now has four extended dimensions. This includes the human element, which requires extra attention and knowledge from students. Students will struggle to understand this dimension, and will

need understand it by utilizing their CT skills in everyday life. Students who master CT skills at high school level will perform better academically in college (Changwong, Sukkamart & Sisan, 2018). CT becomes more important in dealing with this world with its rapid exchange of information (Permana, Hindun, Rofi'ah & Azizah, 2019). By possessing adequate CT skills, students are expected to be better prepared for a future full of demands and challenges.

SE provides hope for parents (Bray, 2006), through expectations about how their children can properly understand school lessons, and how their children can compete and successfully achieve their goals. SE seems to be a phenomenon very dear to Indonesian society, where the thought is formed that as long as the children have taken SE, everything will be fine. However, data show that many children fail to achieve their goals (Shulrufa, Keuskampb & Brakea, 2010), such as entering their desired college university, even though they have taken special classes in their SE. This is certainly very disappointing since SE has become the main hope for parents in their children's education. However, SE has not responded to the challenges (Zhang, 2018) regarding competition among students nationally and the difficulty level to enter tertiary college or university.

This research's preliminary study showed that out of 193 students in 3 schools in Surakarta, Indonesia, 104 have taken SE. This means that more than 53% of the sample have taken SE. From the preliminary observations, economic factors did not prevent participants from taking SE. Indonesia currently has interesting SE options, including affordable classes for the society which are not significantly difference than the expensive ones. That way, anyone is able to somehow take SE. Surprisingly, becoming a private tutor is not a new tide in Indonesia. Many pre-service student teachers join this tide and cannot be blamed for this. They seize this opportunity as a way to learn and improve their skills in the educational field. Furthermore, in doing so, they can make an income or supplement it. The large number of pre-service student teachers in Indonesia who become tutors has implications for SE development. Pre-service student teachers acquire practice to hone didactic knowledge, better understand the complexity of subject matter, and increase their experience in interacting with students. Being a private tutor also offers them various other benefits. Moreover, this kind of exercise has proven to be mind-opening for the pre-service student teachers towards learning activities (Pop, 2015).

Although there have been many discrepancies by consumers regarding SE (Javadi & Kazemirad, 2020), it is still beneficial, especially for the tutors. Therefore, SE has inevitably developed into businesses and even franchises (Aurini & Davies, 2004). SE can be divided into three groups, as pictured in the inverted pyramid in Figure 1. The upper group is the SE that dominates the market (based on student questionnaires in three schools in Surakarta, Indonesia). We labeled the upper group as the bona fide SE institutions, followed by the middle group that does not have many branches but has great trust from society. The last is the private classes that are usually run personally by private tutors.

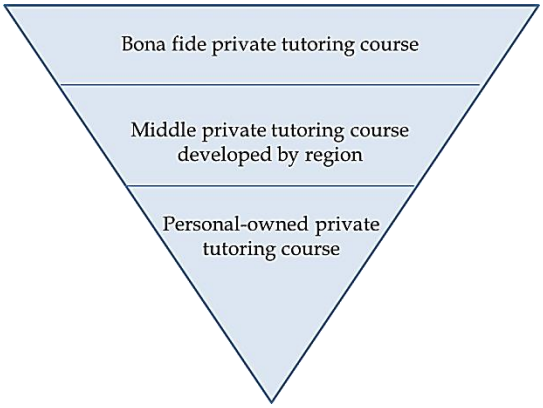

Figure 1: The group of SE in Indonesia divided by its market

The bona fide SE institution has more advantages related to professionalism compared to the other two SE groups. It is highly trusted by the public because of its professionalism in maintaining service quality with good management and marketing. This group also makes promises in their advertising, such as students achieving a high score at school, students being accepted at great universities, and even applying a cashback system on payments. From this can be judged that this group is truly business oriented. However, do these promises really come true? In addition, in the 21st century, students should have good cognitive abilities and several important skills such as communication, creativity, collaboration, CT, literacy, and many other skills (Fisk, 2017). How can SE meet these needs? SE is believed to help students improve their cognitive abilities, improve their learning development diagrams, and pass school tests. However, SE still ignores the essence of real learning. SE is still focused on students' final results while learning. Meanwhile, every learning skill is obtained through a long and iterative process. SE pays less attention to the elements of the psychology of learning (Zhang, Ma & Wang, 2020), even though this is an important key to learning. Students need not only cognitive ability but also to improve these other skills. In fact, the formation of student skills is snowballing and must be done in each education sector transversely (OECD, 2015). Nevertheless, how does SE provide these needs, especially in CT skills?

2. Research Methodology

2.1 General Background

This research used a mixed methods design, a specific explanatory design that simultaneously combines quantitative and qualitative research (Creswell, 2006). This design was chosen to ensure the data's credibility. This research was performed between the 2nd of September 2019 through the 4th of October 2019 in three different high schools (referred to as X, Y, and Z) in Surakarta, Indonesia.

2.2 Sample

The total number of people involved in this study was 217. These can be divided into 9 instrument validators, 193 testees (which include students who filled in the questionnaire and followed the interview), and 15 tutors at SE institutions.

Of the 193 testees, 66 were from School X, 64 from School Y, and 63 from School Z. Meanwhile, from the 193 testees, 45 testees (SE takers) filled questionnaires, every 15 testees represented each school (School X, School Y, and School Z). These groups of 15 testees were also divided into three smaller groups of five each from each respective SE institution (bona fide, middle, and private-personal). School X is one of the most favored schools in Surakarta; most of its students possess more initial knowledge compared to students in other schools. The learning model used at School X is also diverse, including the use of e-learning facilities. Therefore, tests using a computerized testlet at this school are very suitable. On the other hand, School Y has relatively many applicants because this school is a middle-class school based on its difficulty level of entry. Therefore, this school is suitable as choice for almost everyone in the city. This school is located in the city center and has adequate computer laboratory facilities, although it is not as accomplished as School X in e-learning usage. In contrast, School Z is a well-known private school with diverse kinds of students. This school is also located in the city center and has computer laboratory facilities, though rarely used. Thus, their e-learning experience is not good enough compared to Schools X and Y. These kinds of facilities are believed to be able to encourage students to make more efforts to develop their CT (Irawan, Rahardjo & Sarwanto, 2017).

2.3 Instruments and Procedures

The computerized testlet instrument was used to obtain quantitative data about the students' CT skills. A testlet is a test instrument that has several sets of questions. Each set consists of one stimulus, which is used in three different items below it. The stimulus in question can be in the form of reading passages, diagrams, pictures, or other appropriate contexts (Frey, Seitz & Brandt, 2016). The computerized testlet is a testlet-type instrument presented in computerized form. The testees consisted of both SE takers and non-takers. The results would show the difference between SE takers and non-takers based on their CT skills as assessed by the instrument. Computer-based assessment in the learning environment also advantages students to escalate their CT skills (Kopotun, Durdynets, Teremtsova, Markina & Prisnyakova, 2020), which is why the computerized testlet was used.

The quantitative data were analyzed using Rasch modeling. Questionnaires and interviews were used to obtain qualitative data about students' tendencies in their SE. Interviews were also performed with SE tutors to hear their voices about the SE field in which they worked. After the quantitative research had been conducted, the qualitative data were collected and then interpreted to produce a result (Figure 2).

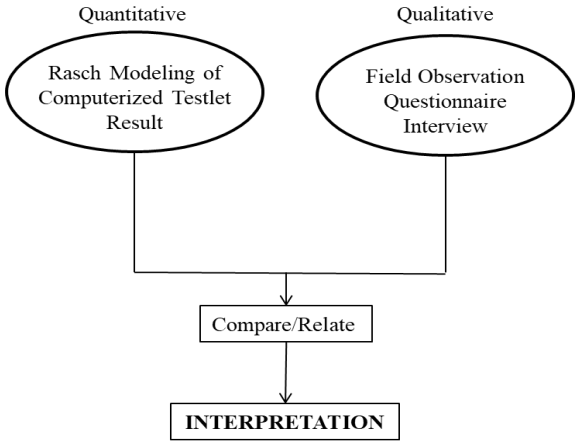

Figure 2: Explanatory design used in this research

The aim of the quantitative data was to reveal students' cognitive abilities. This data were collected using 30 items of a computerized testlet instrument to measure the CT skills domain based on Facione's (2015) on the hydrocarbon topic. The hydrocarbon topic was chosen because it is one of the first yet new materials to be obtained at the 11th grade. It is a difficult topic based on national exams and daily test results (Educational Assessment Center, 2019). Nine validators validated the instrument through a focus group discussion. The validators were asked to complete the validation sheet about the fitness of stems, items, and answer choices based on its proper context, subject knowledge, clarity, spelling, and language (Fahmina, Masykuri, Ramadhani & Yamtinah, 2019). The validators were also experts in educational evaluation and Chemistry learning. The validation result was calculated using Aiken's validity (1985), which showed a validity value of more than 0.75 for each stem, item, and answer choice, as presented in Table 1.

Table 1: Aiken's validity value of the quantitative research instrument

Item	Validity Value		
	Stem	Question	Answer Choices
1	0.85	0.85	0.89
2		0.96	1.00
3		0.81	1.00
4	0.96	0.93	0.85
5		0.81	0.89
6		0.96	0.89
7	0.81	0.96	0.89
8		0.89	0.96
9		0.96	0.89
10	0.96	0.89	0.85
11		0.85	0.96
12		0.96	0.96
13	0.96	0.89	1.00
14		0.93	1.00
15		0.85	0.85
16	0.85	0.89	1.00
17		0.96	1.00

Item	Validity Value		
	Stem	Question	Answer Choices
18		0.89	1.00
19	0.85	0.96	1.00
20		0.89	0.81
21		0.96	1.00
22	0.96	0.89	0.89
23		0.89	0.96
24		0.96	0.85
25	0.93	0.93	0.85
26		0.89	0.93
27		0.89	1.00
28	0.85	0.89	0.93
29		0.96	0.85
30		0.93	0.93

On the other hand, the qualitative data were obtained through an open-ended questionnaire combined with an interview about students' satisfaction on SE and an interview addressed to tutors about the learning strategies on running the SE.

2.4 Data Analysis

Quantitative data were analyzed using the Rasch Model to evince SE takers and non-takers' mapping with their CT skills assessed. That way, we could determine how the testees responded to each item and how they functioned appropriately as a measuring instrument for each group of testees. We could further determine through this data whether testees' response patterns contained specific profile data that cannot be displayed by other measurement tools. The data are about the testees' responses to the items that indicate their CT skills. The quantitative data analyzed are dichotomous data and were analyzed using Winstep 3.73 software with outputs, including Wright maps, DIF plots, and summary statistics of items. The Wright map showed the relationship between the testees' CT skills with item-difficulty levels, so that we could specifically explain how each testee's CT skills relate to each item's difficulty level. The Wright map also allowed seeing outlier data and how each group and person groups were distributed. Through the person measure, we could see what each testee's ability to answer questions looked like, how many total questions could be answered, and in which group the testee was in. Moreover, we could also re-check the testee's demographic data. The DIF plot showed the presence or absence of item discrimination against certain groups of testees. DIF plots presented in graphical form make it easier for readers to analyze item discrimination in certain students' groups. The summary statistics of an item aim to show basic data from the instrument's statistical calculations. This data indicate the reliability of the items, standard deviation (SD), separation, the size of the logit (shown in the column measure), mean square (MNSQ), and Z-standard value (ZSTD). The amount of separation can indicate item level or H, that is by calculating the following:

$$H = \frac{[(4\ x\ separation) + 1]}{3}$$

Qualitative data were analyzed using qualitative analysis based on Miles, Huberman and Saldana (2014), through data collection, data display, data reduction, and conclusion making to obtain compact data following qualitative data collection objectives. Data collection was done using questionnaires addressed to SE takers (students) to determine their level of satisfaction in their experience of joining SE. Questionnaires were followed up with a brief interview to confirm testees' reasons. Interviews were also conducted with SE tutors to determine their perspectives on SE. Data display was conducted to facilitate in archiving and filing. The analysis to be carried out thus became systematic. Data reduction functions to avoid non-essential residual data obtained during the data collection process. The presentation of conclusions was done to display the primary data needed.

3. Research Results

The reliability of the instruments showed good results in the three schools sampled. The instruments' summary statistics aim to display the conclusion of the statistical calculation of the item set. Through this statistical calculation, we can see the average of the item, the SD, the size of the logit (measure) MNSQ, ZSTD, separation, reliability, and the resulting Cronbach alpha value. The SD of the model is 2.00, which is slightly different from the real SD of 1.98. The real item separation is 7.64, which indicates the broader item level (H) that can applies to all students. The reliability of the item, which is 0.98 for both the real and model calculations, indicates that the instrument is acceptable and reliable for performing CT-skills measurements. Besides the reliability of the instrument, it is also important to determine the bias index of the instrument. This analysis is used to determine whether the items affect or disadvantage some groups of people or not. The DIF plot in Figure 3 presents the bias index.

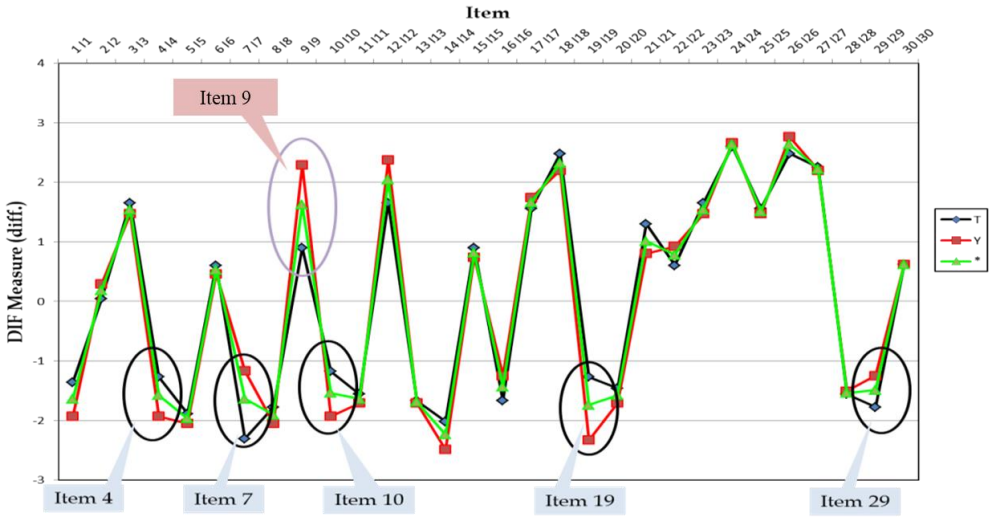

Figure 3: Person DIF plot of DIF measure vs. item

There were several bias effects on the total items, with the bias being quite interesting. The bias was divided into two groups: one of items advantaging the

SE takers (Y – red line) and another with items advantaging the SE non-takers (T – blue line). A summary of the items is presented in Table 2.

Table 2: The context of the items that exhibit bias for SE takers and non-takers

Advantaged Group	Item Number	Context of Items
SE Takers	9	Chemical displacement reaction
SE Non-takers	4	The structural formula (determination of the primary carbon atom)
	7	The structural formula (determination of alkanes, alkenes, and alkynes)
	10	The structural formula (determination of the monomer structure formula from a polymer)
	19	Basic knowledge about the Greenhouse Effect (primary gas that caused the Greenhouse Effect)
	29	Basic knowledge about hydrocarbon combustion

Table 2 shows that SE takers were mostly disadvantaged by the items related to the structural formula for hydrocarbons and the basic knowledge of this material (item number 4, 7, and 10). The SE takers had much trouble with this type of problem, but the SE non-takers on average could handle this type of problem and were better at it than the SE takers. Meanwhile, the average SE non-taker had difficulty in the chemical displacement reaction (item number 9). This study does not, however, discuss and compare the school curriculum with the curriculum of SE institutions, which might be the gap that causes this kind of situation to occur. The questionnaire data show that 104 or 53% of the 193 participating students attended SE. The distribution of their SE group is illustrated in Figure 4.

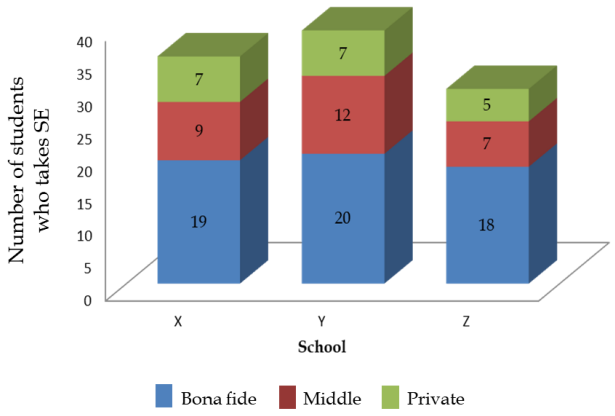

Figure 4: Distribution of SE institution attended by testees at School X, Y, and Z

Figure 4 shows that the number of participating students attending bona fide SE institutions was higher than for middle and private-personal institutions, which occupied a lower preference.

3.1 The Influence of Shadow Education on Students' Ability in Critical Thinking Skills

How does SE affect students' CT skills? The results of the quantitative analysis of students' cognitive abilities are shown on a Wright map depicted in Figure 5. In this figure, two columns are shown. The left shows the distribution of testees, and the right shows the distribution of items. Point 0 shows the median of data distribution. The positive direction shows that the testee has high CT skills because they could answer all the questions below. In the item section, the higher the item, the more difficult it is (Sumintono & Widhiarso, 2015). This also applies for items under 0 (negative direction). Thus, the lower the testee, the lower their CT skills; and the lower the item, the easier it is.

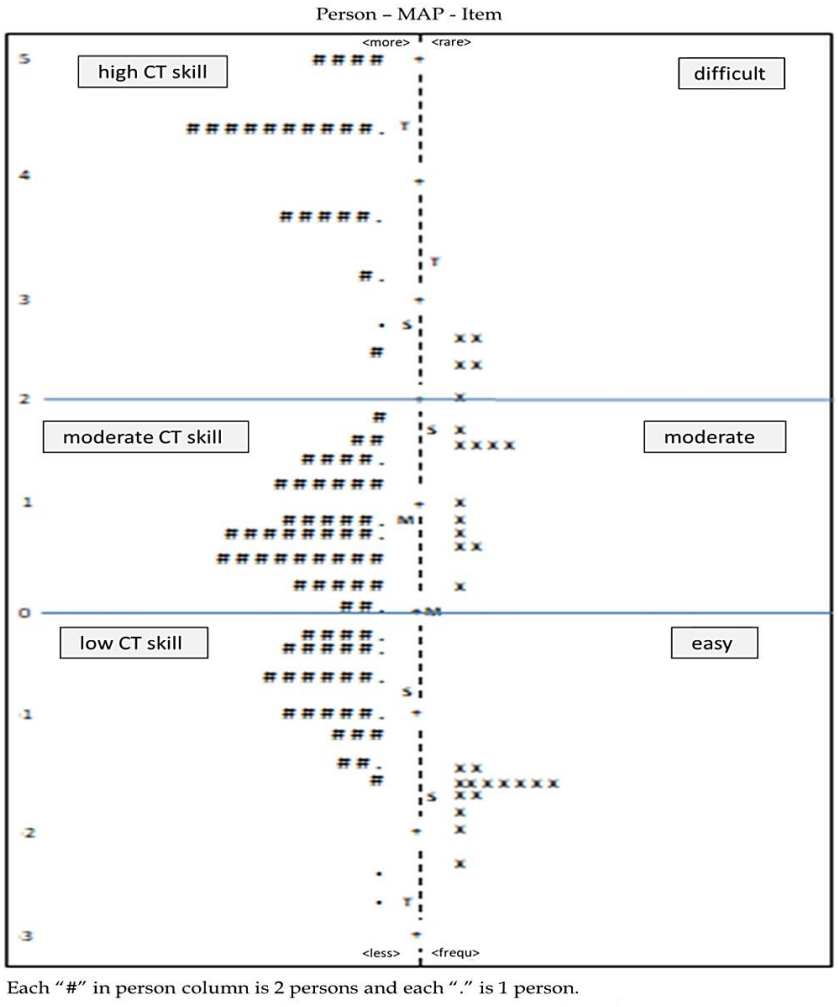

Figure 5: Wright map of person (testee) and item distribution

Eight testees fell above the T point (2SD) of the person column, which means that these eight had the highest CT skills compared with the other testees. However, they were also outliers because they had an abnormal response pattern due to the item set. The outliers are composed of six SE takers and two SE non-takers (Table 3), which means that 25% of the smartest critical thinkers were non-takers of SE. This data show that it is still possible for some students who do not take private courses or SE to obtain high CT skills as expected.

Table 3: Outliers identification

School	Quantity	
	SE Takers	SE Non-takers
X	4	2
Y	2	0

According to Rasch's item level (H level), the outliers are not the only ones to be concerned with. There are also other high critical thinker students and two other groups to consider (Figure 6). The upper group, who had high CT skills, was composed of 38 students (excluding the outliers), 44.7% of whom were SE takers. The middle group, who had moderate CT skills, was occupied by the most testees, 88 students, of which 57.9% were SE takers. The lower group, who had low CT skills, was occupied by 59 testees, 50.8% of whom were SE takers. A summary of this quantitative data showing the composition of the three groups is shown in Figure 6.

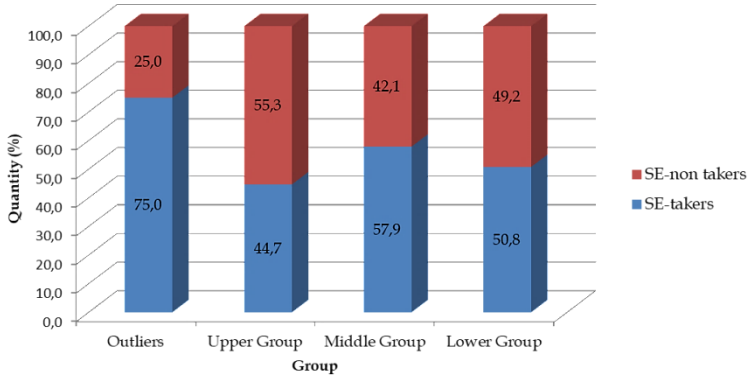

Figure 6: Distribution of SE takers and non-takers in each group of testees

The average of SE takers in all the groups combined (including the outliers) was 57.1%. More data are needed, however, to make a final assumption, as it would be unwise to judge that students will become smarter if they take a private course or vice versa. Taking a private course or not does not necessarily mean that students will succeed in their studies. This concurs with research by Subedi (2008), that relying on private courses will not simply ensure success in students' work on problems given to them. In other words, taking a private course may not guarantee the welfare of students' cognitive abilities at school. The qualitative data of the SE-taker questionnaire are presented in Table 4.

Table 4: Questionnaire results of SE-taker testees about their SE quality

Question	Response	Percentage (%)
1. Reason for taking private course/shadow education	Parents' advice	60
	Self-motivation	40
2. Satisfaction with the private tutor's teaching method	Satisfied	66.7
	So-so	22.2
	Unsatisfied	11.1
3. Private tutor's teaching style	Interesting	84.4
	Boring	15.6

4. Students' preference for studying with …	Teacher	55.6
	Private tutor	44.4
5. SE is beneficial to increase students' cognitive score	Agree	71.1
	So-so	15.6
	Disagree	13.3

The reason of taking SE mostly comes from parents. For question 2, the "satisfied" testees mentioned that their tutors had a more flexible teaching approach. Simultaneously, those who were "so-so" and "unsatisfied" enjoyed their tutoring less because they felt less comfortable with their private tutors. Moreover, sometimes, testees' private tutors were confused in explaining a certain topic and sometimes did not repeat what had been explained, leaving the testees unclear on the topic. Testees who answered "interesting" for question 3 stated that their private tutors often utilized videos and various media when teaching. In contrast, those who answered "boring" claimed that their private tutors only utilized conventional teaching methods. For question 4, testees who answered "teacher" felt comfortable when taught by their school's teachers because the teaching was more systematic than that of private tutors. In contrast, those who answered "private tutor" felt that their teachers were confused and boring when teaching in class. Testees who answered "agree" for question 5 experienced the benefits after having attended private courses because they felt that the lessons helped them to understand those topics in which they were left behind or confused about at school. Those who answered "so-so" claimed that what is taught at school sometimes slightly differs from what tutors teach. Those who answered "disagree" claimed that their grades remained low and considered the lesson to be difficult indeed.

Meanwhile, the interview results of SE tutors' perspectives about SE were elicited by several questions posed to them. The results show that the SE curriculum is appropriate for the school curriculum, especially for the cognitive purpose. SE teaching is usually strengthened drill and practice, not teaching for understanding. Moreover, SE's function is limited in that it aims only to help students complete their homework and prepare for testing, but not to widen their skills. The orientation of the SE institution is excessively directed to the business sector. Some of them do not consider the quality they have to provide. This can be detrimental for students and parents.

4. Discussion

The computerized testlet instrument was used to facilitate students' thinking processes while undertaking the test. The privilege of the testlet instrument is the stem (item stimulant). Therefore, a testlet can be used as a stimulus for students' understanding of knowledge (Yamtinah et al., 2019). Utilizing the stimulant as a resource can help students develop their CT (Stephenson & Sadler-McKnight, 2016). CT has a primary role in students' success at the college level and working environment (Rusmansyah, Yuanita, Ibrahim, Isnawati & Prahani, 2019). By utilizing the test, we could discover the testees' responses to CT assessment. The quantitative analysis results showed that of the testees identified as outliers or the smartest critical thinkers, 25% were SE non-takers. Of

the testees with high levels of CT skills (except the outliers), 55.3% were SE non-takers. The upper group, including the outliers, had an equal percentage of 50% each of SE takers and non-takers, which shows that they do not differ significantly. SE does thus not significantly improve the CT of students who are already smart at school and have sufficient initial CT skills. If we look at the middle and lower groups, there were exciting and significant differences between SE takers and non-takers. The middle class was constituted with more SE takers (57.9%). This shows that SE has a significant influence on helping students acquire good CT skills for moderate cognitive ability at school. There was also a common border with the lower group's results, with 50.8% of testees being SE takers. The learning process provided by SE is predominantly about memory questions. This makes SE less effective for students who are already smart. However, for students with moderate and low intelligence levels, SE can help them remember information. SE functions like these must nonetheless be improved. In addition, SE must catch up on what the world wants, what developments the world needs. Punjabi (2019) also mentioned that SE must be more focused and serious on student development in active learning, and not stuck in facilitating passive learning.

The qualitative data results indicate that qualified tutors at SE institutions have an important role in student learning activities and satisfaction. Tutors who are fun and have good teaching skills are in higher demand by students, bearing in mind that their goal with SE is to level up students' abilities at school. These reasons are supported by Bray (2006), who indicated that the pattern of learning instructions of SE is mostly different from formal schools. Moreover, Yung (2019) proposed that SE mostly use a drill and practice approach. SE in countries outside Indonesia has won the public's trust that the teaching methods used are more adaptable than those of formal schools which tend to be more conservative (Zhang & Bray, 2020). This contrasts with SE in Indonesia, which is less adaptable, as evidenced by this study's results. Moreover, interviews with the 15 SE tutors showed that, in general, the SE curriculum has adjusted to the school curriculum even though it is still limited to cognitive needs. Thus, SE is not yet entirely focused on the needs of other skills besides cognitive abilities. Students also need to be trained to have the skills needed in this 21st century. Besides, SE's primary function is to get students used to working on problems, not to understand. Consequently, at this point, SE must be accompanied by a self-evaluation, remembering that students need time to compete, not only for answering questions, but also for doing things like applying the concept to solve similar future problems. In addition, being business oriented is not wrong, considering that SE is also an economic business activity. However, this concept must be applied together with quality improvement by evaluating and reorienting SE objectives.

Through this research, it is hoped that society can make a good assessment of SE selection, given the many important factors that need to be considered. Moreover, society must be sure and aware of the real purpose of registering children at a private-course institution (Chui, 2016) to achieve cognitive scores and other skills needed. Further research on the similarities and differences

between the school curriculum and SE curriculum also needs to be done to find important gaps to help SE perform its function as appropriate supplementary tutoring. SE developers must also begin to improve their learning curriculum so that SE can also become a formal education companion facility as a learning center that is aware of the learning process in forming skills, especially CT skills. That way, SE becomes more relevant as supplementary education that can help students with cognitive abilities and CT skills needed for education later on. Moreover, in developing student CT, educational stakeholders, including those in SE, must work in an upright manner (Almeida & Franco, 2011).

5. Conclusions

Based on the quantitative and qualitative results, supplementary private tutoring or SE cannot significantly improve students' CT skills for those who initially have good cognitive ability at school. However, joining SE may be good for those who have moderate cognitive ability at school. It does not mean that this private course is not useful at all, but parents must be selective and critical in choosing a private course for their child. Knowledge and understanding of parents ultimately become an important key to the success of students at school. Besides, students must succeed in their studies and compete globally by possessing the skills needed in the 21st century, especially CT skills. This research has shown several important points that must be considered by the public in determining the choice of SE that suits the needs of students in this 21st century. In addition, educational researchers are also expected to conduct further research to uncover the gap between the school curriculum and the SE curriculum so that SE can improve and adjust its objectives to appropriately supplement teaching. Moreover, after seeing the prospect of SE development, which is increasingly in demand by the society, it is necessary to improve SE's quality and objectives. That way, SE will help students remember lessons and be a companion for formal schools in improving student skills.

6. Limitation of This Study

This study has several limitations. Firstly, the topic used in the computerized testlet instrument was hydrocarbon material. In addition, on the bias of some items, we have not conducted further research on the possibility of explicit curriculum differences between the school curriculum and SE-institution curriculum.

7. Acknowledgments

We would like to show gratitude to the Institute of Research and Community Services of Universitas Sebelas Maret, which supported us in conducting this research through funding on Master's Thesis Research contract number 516/UN27.21/PP/2019.

8. References

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

Almeida, L. da S., & Franco, A. H. R. (2011). Critical thinking: Its relevance for education in a shifting society. Retrieved December 19, 2020, from

http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0254-92472011000100007&lng=es&tlng=en

Aurini, J., & Davies, S. (2004). The transformation of private tutoring: Education in a franchise form. *The Canadian Journal of Sociology*, *29*(3), 419–438. https://doi.org/10.1353/cjs.2004.0031

Bray, M. (2006). Private supplementary tutoring: Comparative perspectives on patterns and implications. *Compare: A Journal of Comparative and International Education*, *36*(4), 515–530. https://doi.org/10.1080/03057920601024974

Bray, M. (2013). Benefits and tensions of shadow education: Comparative perspectives on the roles and impact of private supplementary tutoring in the lives of Hong Kong students. *Journal of International and Comparative Education*, *2*(1), 18–30. https://doi.org/10.14425/00.45.72

Bray, M. (2014). The impact of shadow education on student academic achievement: Why the research is inconclusive and what can be done about it. *Asia Pacific Educational Review*, *15*, 381–389. https://doi.org/10.1007/s12564-014-9326-9

Changwong, K., Sukkamart, A., & Sisan, B. (2018). Critical thinking skill development: Analysis of a new learning management model for Thai high schools. *Journal of International Studies*, *11*(2), 37–48. https://doi.org/10.14254/2071- 8330.2018/11-2/3

Chui, M. M. (2016). Private supplementary tutoring: Motivations and effects. *Journal of Education and Practice*, *7*(27), 195–198. Retrieved from https://eric.ed.gov/?id=EJ1115807

Creswell, J. W. (2006). Choosing a mixed method design. In J. W. Creswell & V. L. P. Clark (Eds.), *Designing and conducting mixed methods research* (pp. 58–89). Thousand Oaks, CA: Sage.

Directorate General of Early Childhood, Primary, and Secondary Education (2020). Basic data of primary and secondary education [Data pokok pendidikan dasar dan menengah]. Retrieved December 15, 2020, from https://dapo.dikdasmen.kemdikbud.go.id/sp

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. https://doi.org/dx.doi.org/10.1016/j.tsc.2013.12.004

Educational Assessment Center. (2019). Students' mastery of National Exam [Penguasaan materi Ujian Nasional]. Retrieved December 15, 2020, from https://hasilun.puspendik.kemdikbud.go.id/#2019!sma!daya_serap!03&02&999!a&05&T&T&1&!1!&

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Hermosa Beach, CA: Insight Assessment.

Fahmina, S. S., Masykuri, M., Ramadhani, D. G., & Yamtinah, S. (2019). Content validity uses Rasch Model on computerized testlet instrument to measure chemical literacy capabilities. In *AIP Conference Proceedings 2194* (Vol. 020023), July 27, 2019. Surakarta: AIP Publishing. https://doi.org/10.1063/1.5139755

Fisk, P. (2017). Education 4.0 … the future of learning will be dramatically different, in school and throughout life. Retrieved August 2, 2019, from http://www.thegeniusworks.com/2017/01/future-education-young-everyone-taught-together

Frey, A., Seitz, N.-N., & Brandt, S. (2016). Testlet-based multidimensional adaptive testing. *Frontiers in Psychology*, *7*(1758). https://doi.org/10.3389/fpsyg.2016.01758

Irawan, T. A., Rahardjo, S. B., & Sarwanto. (2017). Analysis of secondary school students' critical thinking skill in learning energy in living system. *Pancaran Pendidikan*,

6(4), 1–8. https://doi.org/10.25037/pancaran.v6i3.78

Javadi, Y., & Kazemirad, F. (2020). Worldwide shadow education epidemic and its move toward shadow curriculum. *Journal of Language Teaching and Research*, *11*(2), 212–220. https://doi.org/10.17507/jltr.1102.09

Kopotun, I. M., Durdynets, M. Y., Teremtsova, N. V., Markina, L. L., & Prisnyakova, M. L. (2020). The use of smart technologies in the professional training of students of the Law Departments for the development of their critical thinking. *International Journal of Learning, Teaching and Educational Research*, *19*(3), 174–187. https://doi.org/10.26803/ijlter.19.3.10

Krishnakumar, J., & Nogales, R. (2020). Education, skills and a good job: A multidimensional econometric analysis. *World Development*, *128*(104842), 1–17. https://doi.org/10.1016/j.worlddev.2019.104842

Mahaffy, P. (2004). The future shape of chemistry education. *Chemistry Education: Research and Practice*, *5*(3), 229–245. https://doi.org/10.1039/B4RP90026J

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative Data Analysis* (3rd ed.). New York: Sage.

Mulyani, E., Widiastuti, A., & Supriyanto. (2019). Developing a model for evaluating learning outcomes of an entrepreneurship course. *International Journal of Learning, Teaching and Educational Research*, *18*(11), 53–69. https://doi.org/10.26803/ijlter.18.11.4

Nam, Y., & Chan, K. (2019). The roles of mainstream schooling and shadow education in english language teaching: A case study in Hong Kong. *Education Journal*, *8*(1), 16–26. https://doi.org/10.11648/j.edu.20190801.13

OECD. (2015). Learning contexts that drive skill formation. In *Skills for Social Progress: The Power of Social and Emotional Skills* (pp. 73–94). Paris: OECD Publishing. https://doi.org/10.1787/9789264226159-en

OECD. (2018). Indonesia student performance (PISA 2018). Retrieved March 5, 2020, from https://gpseducation.oecd.org/CountryProfile?primaryCountry=IDN&treshold=10&topic=PI

Permana, T. I., Hindun, I., Rofi'ah, N. L., & Azizah, A. S. N. (2019). Critical thinking skills: Its correlation with academic ability, mastering concepts, and analytical skill of undergraduate students in botany class. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *5*(1), 1–8. https://doi.org/10.22219/jpbi.v5i1.7626

Pop, R. (2015). Understanding pre-service trainees' perceptions of their teacher training experience. *Procedia – Social and Behavioral Sciences*, *209*, 378–382. https://doi.org/10.1016/j.sbspro.2015.11.207

Punjabi, S. (2019). Is shadow education becoming the "new" formal? Effects of pedagogical approaches of IIT-JEE coaching on school education in the city of Delhi. *Contemporary Education Dialogue*, *17*(1), 14–44. https://doi.org/10.1177/0973184919885485

Rusmansyah, R., Yuanita, L., Ibrahim, M., Isnawati, I., & Prahani, B. K. (2019). Innovative chemistry learning model: Improving the critical thinking skill and self-efficacy of pre-service chemistry teachers. *Journal of Technology and Science Education*, *9*(1), 59–76. https://doi.org/10.3926/jotse.555

Shulrufa, B., Keuskampb, D., & Brakea, D. (2010). The impact of course-taking on academic achievements: A systematic review and meta analysis. *Procedia Social and Behavioral Sciences*, *2*, 3401–3406. https://doi.org/10.1016/j.sbspro.2010.03.523

Stephenson, N. S., & Sadler-McKnight, N. P. (2016). Developing critical thinking skills using the science writing heuristic in the chemistry laboratory. *Chemistry*

Education Research and Practice, *17*(1), 72–79. https://doi.org/10.1039/C5RP00102A

Subedi, K. R. (2008). Shadow education: A role of private tutoring in learning. *International Journal of Humanities and Social Sciences*, *1*(2), 29–42. Retrieved from https://eric.ed.gov/?q=Shadow+Education%3A+A+Role+of+Private+Tutoring+in+Learning&id=ED591992

Sumintono, B., & Widhiarso, W. (2015). *Rasch modeling application in education assessment [Aplikasi pemodelan Rasch pada assesment pendidikan]*. Cimahi: Trim Komunikata.

Woldeamanuel, M. M., Atagana, H., & Engida, T. (2014). What makes chemistry difficult? *African Journal of Chemistry Education*, *4*(2), 31–43. Retrieved from https://www.ajol.info/index.php/ajce/article/view/104070

Worldometer. (2020). Indonesia population. Retrieved December 15, 2020, from https://www.worldometers.info/world-population/indonesia-population/

Yamtinah, S., Saputro, S., Mulyani, S., Ulfa, M., Lutviana, E., & Shidiq, A. S. (2019). Do students have enough scientific literacy? A computerized testlet instrument for measuring students' scientific literacy. In *AIP Conference Proceedings 2194* (p. 020143), July 27, 2019. Surakarta: AIP Publishing. https://doi.org/10.1063/1.5139875

Yung, K. W.-H. (2019). Learning, teaching, and researching in shadow education in Hong Kong: An autobiographical narrative inquiry. *ECNU Review of Education*, *2*(1), 64–76. https://doi.org/10.1177/2096531119840871

Zhang, W., & Bray, M. (2020). Comparative research on shadow education: Achievements, challenges, and the agenda ahead. *European Journal of Education*, *55*, 322–341. https://doi.org/10.1111/ejed.12413

Zhang, Y., Ma, X., & Wang, L. (2020). The determinants of private tutoring participation for mathematics in China: Focusing on the role of student metacognition. *Frontiers in Psychology*, *11*(603), 1–10. https://doi.org/10.3389/fpsyg.2020.00603

Zhang, Y. (2018). Private tutoring, students' cognitive ability and school engagement, and the formal schooling context: Evidence from middle school students in China. *Chinese Journal of Sociology*, *4*(2), 277–298. https://doi.org/10.1177/2057150X18764859

Appendixes

Appendix I: The questionnaire form and interview questions used to reveal students' satisfaction with shadow education

1. What is your reason for taking a private course/ shadow education?
   □ Parents' advice
   □ Self-motivation
   Elaboration from the interview: ……
2. Are you satisfied with how your instructors teach you in your private course/ shadow education?
   □ Satisfied
   □ So-so
   □ Unsatisfied
   Elaboration from the interview: ……
3. How is your private tutors' teaching style in your private course/ shadow education?
   □ Interesting
   □ Boring
   Elaboration from the interview: ……
4. Which one do you prefer, study with your school's teachers or your private course/ shadow education's instructors?
   □ Teacher
   □ Private tutor
   Elaboration from the interview: ……
5. Do you feel the significant benefit (especially on increasing your cognitive score) after taking private course/ shadow education?
   □ Agree
   □ So-so
   □ Disagree
   Elaboration from the interview: ……

Appendix II: The interview questions used to reveal private tutors' perspectives about shadow education in which they worked

1. How do you see the fit between the shadow education curriculum and the school curriculum?
2. How do you describe the main characteristic of shadow education compared to the school's teaching purpose?
3. How do you think about the shadow education function?
4. How do you think about the shadow education orientation, considering there are more and more new institutions emerging to offer shadow education?

Appendix III: The summary statistic of items analyzed by Winstep 3.73

	Total Score	Count	Measure (Logit)	OUTFIT	
				MNSQ	ZSTD
Item					
MEAN	120.30	193.00	0.00	0.96	0.10
SD	45.80	0.00	1.68	0.19	0.60
MAX	176.00	193.00	2.63	1.36	1.60
MIN	52.00	193.00	-2.24	0.58	-1.00
REAL TRUE SD = 0.22		SEPARATION = 7.64		Item RELIABILITY = 0.98	
MODEL TRUE SD = 0.21		SEPARATION = 7.73		Item RELIABILITY = 0.98	
S.E. OF ITEM MEAN = 0.31					

Appendix IV: The summary statistic of persons analyzed by Winstep 3.73

	Total Score	Count	Measure (Logit)	OUTFIT	
				MNSQ	ZSTD
Item					
MEAN	18.70	30.00	1.10		
SD	6.50	0.00	1.99		
MAX	30.00	30.00	5.72		
MIN	4.00	30.00	-2.72	0.24	-2.70
REAL TRUE SD = 0.72		SEPARATION = 2.58		Item RELIABILITY = 0.87	
MODEL TRUE SD = 0.70		SEPARATION = 2.68		Item RELIABILITY = 0.88	
S.E. OF ITEM MEAN = 0.14					
CRONBACH-ALPHA (KR-20) RELIABILITY = 0.90					