Effectiveness of Higher Order Thinking Skills-based Test Instruments in Science Learning in Indonesia: A Meta-analysis

Tomi Apra Santosa1*, Siska Angreni1, Rona Taula Sari1, Festiyed2, Yerimadesi2, Yuni Ahda2, Heffi Alberida2, Fitri Arsih2

1 Doctor of Science Education, FMIPA, Padang State University, Padang, Indonesia.
2 Postgraduate Lecturer, FMIPA, Padang State University, Padang, Indonesia.

Received: December 19, 2023
Revised: January 5, 2024
Accepted: May 31, 2024
Published: May 31, 2024

Corresponding Author:
Tomi Apra Santosa
santosa2021@yahoo.com

DOI: 10.29303/jppipa.v10i5.6654

© 2024 The Authors. This open access article is distributed under a (CC-BY License)

Abstract: This research is motivated by the many applications of Higher Order Thinking Skills-based test instruments in science learning. However, it has not been found how effective the instrument is in science learning. The purpose of this study was to determine the effectiveness of Higher Order Thinking Skills-based test instruments in science learning. This type of research is a meta-analysis. Data sources come from the databases of Google Scholar, Hindawi, ScienceDirect, Wiley, Taylor of Francis, Sage Journal, Springer, ProQuest and ERIC. Search keywords are test instruments, test instruments based on Higher Order Thinking Skills in science learning, and higher order thinking skills. This study analyzed 20 articles published in 2020-2023 used in the meta-analysis. Data analysis techniques with the help of JSAP software. The results concluded that the summary effect size value was 1.24 (very high effect size). This finding shows that the use of test instruments based on Higher Order Thinking Skills is effectively applied in science learning. Furthermore, the research provides higher order thinking skills-based test instruments in science learning in Indonesia and provides important information for the application of higher order thinking skills-based test instruments in the future.

Keywords: Effect size; Higher order thinking skills; Science learning; Test instrument

Introduction

Natural Science is a branch of science that studies natural phenomena (Uluçınar, 2023; He et al., 2020). Science learning trains students to develop logical and scientific thinking skills in learning (Suendarti & Virgana, 2022; Angganing et al., 2022; Aydede, 2022). In addition, in the science learning process students not only understand a concept but pressure students to have critical, careful and meticulous thinking skills (Sumiantari et al., 2019; Baysal et al., 2023; Kim & Alghamdi, 2023). Not only that, science learning helps students learn to find out about the universe through scientific observations and procedures (Puspita et al., 2023; Bantaokul & Poliyiem, 2022; Wang & Tsai, 2019; Shapiro, 2015).

But in fact, science learning in schools is not in accordance with the current independent curriculum (Azzahra et al., 2023; Nuryani et al., 2023). Science learning is still teacher-centered so that students find it difficult to understand the subject matter (Fahrezi et al., 2020). In addition, the quality of science learning in Indonesia is classified as low (Yakob et al., 2021). This can be seen from the results of research by the 2018 Programme for International Student Assessment (PISA) conducted by the OECD showing the ability of Indonesian students in science literacy to obtain a score of 396 ranked 71 out of 78 countries (Hariyadi et al., 2023;...
Rahman et al., 2023; Utomo et al., 2023; Zulyusri et al., 2023; Ichsan et al., 2023). Teachers use assessment sheets that are not in accordance with the curriculum so that learning objectives are not achieved (Puspita et al., 2023). In addition, in science learning, students have low problem-solving, critical and creative thinking skills (Alomery, 2022; Yustiana et al., 2022; Dwi et al., 2022). Teacher-developed instruments are not suitable for measuring students' higher-order thinking skills in science learning.

For one of the efforts to develop students' abilities in science learning, develop appropriate numeracy test instruments to measure student learning ability. Instrumen numeracy is a tool used to measure students' learning ability (Divayana et al., 2021; Ramadhan & Sumiharsono, 2020). Effective numeracy test instruments are developed in science learning (Qisthi et al., 2023). Numeracy instruments based on higher order thinking skills are widely developed in science learning. Previous research on the application of numeracy instruments effectively develops students' mathematical skills and problem-solving abilities in learning (Purnomo et al., 2022). Development of effective numeracy instruments to improve students' science literacy skills (Sa'dijah et al., 2023).

Furthermore, higher order thinking skill-based test instruments can measure students' higher-order thinking skills in learning. Higher order thinking skill-based test instruments are very important for teachers to develop in science learning so that students have higher-order thinking skills in learning (Rahayu et al., 2023). Quality test instruments can provide the right results in measuring students' cognitive abilities in learning. Research by Apriani et al. (2023) test instruments based on higher order thinking skills are effective and valid developed in science learning. However, research by Zulfiana et al. (2023) states that higher order thinking skill-based test interventions are less effective for measuring students' science literacy. The existence of Kesanja gan in this study needs to be seen comprehensivelyye test intervention based on higher order thinking skills in science teaching. In addition, there are no research findings that measure the effect of interumen test sizes based on higher order thinking skills. Based on this, this study aims to determine the effectiveness of Higher Order Thinking Skills-based test instruments in science learning.

Method

Research Design

This research is a type of meta-analysis research. Meta-analysis research is a type of research that collects and analyzes primary research that can be analyzed quantitatively (Martin & Carolina, 2022; Tamur et al., 2020; Razak et al., 2021; Çevik & Bakioğlu, 2022; Aybirdi et al., 2023; Setiawan et al., 2022). Meta-analysis was conducted to determine the effectiveness of higher order thinking skill-based test instruments in science learning. According to Borenstein et al. (2009) The procedure in meta-analysis research consists of 1) establishing research inclusion criteria; 2) Collection of primary data and coding of data to be analyzed; 3) Perform statistical data tests to analyze data can be seen in figure 1.

Eligibility Criteria

In the search for data sources, articles included in the meta-analysis must meet the inclusion criteria, namely research must be published in 2020-2023; research comes from reputable international journals or proceedings indexed by SINTA and Scopus; Research must have an experimental class applying test instruments based on higher order thinking skills while conventional control classes; And research data must be complete to calculate the value of the effect size. So, from the process of searching for data sources, 20 articles that meet the inclusion criteria can be seen (Table 3). The data selection process using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method can be seen in figure 1.

Data Collection

Data collected in this meta-analysis was based on the databases of Google Scholar, Hindawi, ScienceDirect, Wiley, Taylor of Francis, Sage Journal, Springer, ProQuest and ERIC. Search keywords are test instruments, test instruments based on higher order thinking skills in science learning, and higher order thinking skills. The results of data collection from journal databases can be seen in Table 1.

Table 1. Meta-Analysis Data Collection Results

<table>
<thead>
<tr>
<th>Description</th>
<th>Sum</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Publication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>2021</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>2022</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>2023</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary School</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Junior School</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>High School</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>University</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 1. Procedures in meta-analysis
Statistical Analysis

Data analysis in meta-analysis research is calculating the effect size value of each study (Riopel et al., 2020; Hillmayr et al., 2020). The effect size in this study is an index that describes the influence of higher order thinking skills-based test instruments in science learning. According to Borenstein et al. (2009) the steps for statistical analysis in meta-analysis are calculating the effect size value of the primary study, conducting heterogeneity tests and determining estimation models, checking publication bias and calculating p-value values for hypothesis testing. Statistical analysis in this meta-analysis with the help of JSAP software. Furthermore, the criteria for effect size values in this meta-analysis can be seen in Table 2.

Table 2. Effect Size Value Category (Cohen et al., 2007)  
<table>
<thead>
<tr>
<th>Effect Size Value Category</th>
<th>Effect Size Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>0.00 ≤ ES ≤ 0.20</td>
</tr>
<tr>
<td>Medium</td>
<td>0.20 ≤ ES ≤ 0.80</td>
</tr>
<tr>
<td>Large</td>
<td>ES ≥ 0.80</td>
</tr>
</tbody>
</table>

Furthermore, an important thing to note in conducting meta-analyses is checking publication bias (Smith et al., 2023; Schmid et al., 2023). In a study to check publicity bias using funnel plots and Rosenthal Fail Safe N test (Juandi et al., 2022; Puspita et al., 2022). Table 3, the effect size values of the 20 articles analyzed ranged from 0.53 to 2.51. According to the effect size category Cohen et al., (2007) there are five articles with a medium effect size value (25%) and fifteen articles with a high effect size value (75%). Next, conduct heterogeneity tests and determine the estimation model used in calculating the average effect size of 20 articles. The results of the heterogeneity test and determination of the estimation model with random and fixed models can be seen in Table 4.

Result and Discussion

Based on the analysis of 20 articles that meet the inclusion criteria, effect size values and error standards are calculated which can be seen in Table 3.

Table 3. Effect Size and Standard Error 20 Articles
<table>
<thead>
<tr>
<th>Code</th>
<th>Year</th>
<th>Country</th>
<th>Journal Index</th>
<th>Effect Size</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>2022</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.67</td>
<td>0.21</td>
</tr>
<tr>
<td>Study 2</td>
<td>2020</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>1.22</td>
<td>0.34</td>
</tr>
<tr>
<td>Study 3</td>
<td>2020</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>0.96</td>
<td>0.27</td>
</tr>
<tr>
<td>Study 4</td>
<td>2022</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>1.45</td>
<td>0.22</td>
</tr>
<tr>
<td>Study 5</td>
<td>2023</td>
<td>Indonesian</td>
<td>SINTA</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td>Study 6</td>
<td>2021</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.79</td>
<td>0.20</td>
</tr>
<tr>
<td>Study 7</td>
<td>2022</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.53</td>
<td>0.23</td>
</tr>
<tr>
<td>Study 8</td>
<td>2023</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.82</td>
<td>0.31</td>
</tr>
<tr>
<td>Study 9</td>
<td>2022</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>2.51</td>
<td>0.45</td>
</tr>
<tr>
<td>Study 10</td>
<td>2023</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>1.34</td>
<td>0.31</td>
</tr>
<tr>
<td>Study 11</td>
<td>2021</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>0.95</td>
<td>0.35</td>
</tr>
<tr>
<td>Study 12</td>
<td>2021</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.81</td>
<td>0.20</td>
</tr>
<tr>
<td>Study 13</td>
<td>2021</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>0.88</td>
<td>0.22</td>
</tr>
<tr>
<td>Study 14</td>
<td>2022</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>1.04</td>
<td>0.35</td>
</tr>
<tr>
<td>Study 15</td>
<td>2021</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>2.07</td>
<td>0.40</td>
</tr>
<tr>
<td>Study 16</td>
<td>2023</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>1.30</td>
<td>0.28</td>
</tr>
<tr>
<td>Study 17</td>
<td>2022</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.75</td>
<td>0.20</td>
</tr>
<tr>
<td>Study 18</td>
<td>2022</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>1.70</td>
<td>0.36</td>
</tr>
<tr>
<td>Study 19</td>
<td>2023</td>
<td>Indonesian</td>
<td>Scopus</td>
<td>1.19</td>
<td>0.40</td>
</tr>
<tr>
<td>Study 20</td>
<td>2023</td>
<td>Indonesian</td>
<td>Sinta</td>
<td>0.77</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 4. Heterogeneity Test Results with Random and Fixed Models
<table>
<thead>
<tr>
<th>Q</th>
<th>Dr</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>78.169</td>
<td>1</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>214.086</td>
<td>19</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Table 4, explaining the results of the heterogeneity test, obtained a Q value of 214,086 greater than the value of 78,169 and a p value of < 0.001. This finding concludes that the effect size analyzed is heterogeneously distributed. The random effect model is effective for analyzing the mean effect size 20 articles included in the meta-analysis. The next step is to check the publication bias of the 20 articles included in the meta-analysis. Publication bias checking in meta-analyses with funnel plots and Rosenthal Fail Safe N (Bernard et al., 2014;
Suryono et al., 2023; Li & Wang, 2023; Juandi et al., 2021). The results of the publication bias test can be seen in Figure 3.

![Funnel Plot](image)

**Figure 3.** Funnel Plot Standard Error

Based on figure 3, the results of the analysis of 20 effect sizes with funnel plots cannot be concluded whether the form of funnel plot analysis is symmetric or asymmetric. Furthermore, it is necessary to do the Rosenthal Fail Safe N test. The results of the Rosenthal Fail Safe test can be seen in Table 5.

<table>
<thead>
<tr>
<th>File Drawer Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail Safe N</td>
</tr>
<tr>
<td>Rosenthal</td>
</tr>
</tbody>
</table>

Table 5, shows the Rosenthal fail safe N test result of 2203 with sig value 0.50 and p < 0.00. The Rosenthal fail Safer value compared to the value of k = 20 or k = 5k + 10 = 110, then the Rosenthal fail safe value of N > 5k + 10 means that in the analysis of 20 articles included in the meta-analysis data no publication bias was found. Next, calculate the average value of the effect size to test the hypothesis. The results of the mean effect size test can be seen in Table 6.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>ICE Standard Error</th>
<th>Z</th>
<th>P</th>
<th>95% Confidence Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.04</td>
<td>0.090</td>
<td>11.595</td>
<td>&lt; 0.001</td>
<td>0.868</td>
</tr>
</tbody>
</table>

Table 6, Mean Effect Size

Based on Table 6, explaining the mean effect size value with random effect model is obtained (r_{RE} = 1.04; SE = 0.090; z = 11.59; p < 0.001). This finding shows that the use of higher order thinking skill-based test instruments has a significant effect on science learning in Indonesia. The effect size in this is included in the high category. With a confidence level of 95% with a lower limit of 0.868 and an upper limit of 1.221. Therefore, the application of higher order thinking skills-based test instruments is effective and influential in science learning.

This research is in line with Ahmad et al. (2017) stated that higher order thinking skills-based test instruments are effectively used in science learning. This result is supported by Ramadan et al. (2019) a valid higher order thinking skills-based test instrument that can measure students’ higher-order thinking skills in learning. The use of higher order thinking skills-based test instruments is very necessary in science learning because it can develop students’ critical thinking and analytical skills (Walid et al., 2019; Arafat et al., 2021). Furthermore, higher order thinking skills-based test instruments help students solve higher-order thinking problems (Damaianti et al., 2020).

The use of test instruments based on students’ higher order thinking skills helps teachers be more creative in developing assessments in science learning (Suprapto et al., 2020). In addition, in science learning students must have higher-order thinking skills in order to be able to solve problems in learning (Utama et al., 2020; Wilson & Narasuman, 2020). Not only that, test instruments based on higher order thinking skills students learn more creatively (Samritin & Suryanto, 2016). According to Tanujaya (2016), higher order thinking skills-based test instruments are an effective evaluation tool to measure cognitive abilities in science learning. Test instruments are a tool to determine student success in learning science. Higher order thinking skills-based test instruments are an important solution in evaluating the quality of science learning in Indonesia.

**Conclusion**

From this meta-analysis, it can be concluded that the summary effect size value is 1.04 (very high effect size). This finding shows that the use of test instruments based on Higher Order Thinking Skills is effectively applied in science learning. Furthermore, the research provides higher order thinking skills-based test instruments in science learning in Indonesia and provides important information for the application of higher order thinking skills-based test instruments in the future. The use of higher order thinking skills-based test instruments is an important solution for teachers to evaluate science learning in Indonesia. Higher order thinking skills based test instruments encourage
students to have higher-order thinking skills in science learning.

Acknowledgments
In the research, the researcher thanked the doctoral lecturer of Science Education, Padang State University who had contributed and provided guidance and input in completing this research. In addition, we would like to thank the JPPIPA editorial board for accepting and publishing this article.

Author Contributions
In the research, all researchers contributed Tomi Apra Santosa, Siska Anggreni and Rona Taula Sari collecting, selecting and analyzing data statistically and Festiyed, Yerimadesi, Yuhi Ahda, Heffi Alberida and Fitri Arsih provided guidance, input and checked this article.

Funding
This research received no external funding.

Conflicts of Interest
The authors declare no conflict of interest.

References


