P-ISSN: 1411-3031; E-ISSN: 2442-9651
Available online https://journal.uinsi.ac.id/index.php/dinamika_ilmu
doi: http://doi.org/10.21093/di.v23i2.6979
DINAMIKA ILMU, 23(2), December 2023

DINAMIKA ILMU
Journal of Education
Jurnal Pendidikan

Designing Authentic Science Assessment for Pre-Service Science Teachers

Mustafa Tüysüz[1], Ümmüye Nur Tüzün[2], Hayati Çavuş[3]

[1] *Van Yüzüncü Yil University, Education Faculty, Science Education Department, Van, Türkiye, mustafatuysuz@yyu.edu.tr*
[2] *Ministry of National Education, Ankara, Türkiye, u_tuzun@hotmail.com*
[3] *Van Yüzüncü Yıl University, Education Faculty, Information and Computer Technologies Department, Van, Türkiye, hcavus@yyu.edu.tr*

Received 30 June 2023 | Received in revised form 12 August 2023 | Accepted 01 September 2023

APA Citation:
Tüysüz, M., Tüzün, U. N., & Çavuş, H. (2023). Designing Authentic Science Assessment for Pre-Service Science Teachers. *DINAMIKA ILMU, 23(2)*, 261-276.
http://doi.org/10.21093/di.v23i2.6979

Abstract

The current research aimed to design an authentic science assessment domain for pre-service science teachers to enhance their higher thinking skills, such as critical thinking skills, being aware that if pre-service science teachers experienced effective authentic assessment domains, they would be much more skillful in their future classes. The research was conducted with the case study as one of the qualitative research designs. The authentic assessment process was performed on Cunningham engineering design integrated with Toulmin argument pattern components. Worksheets were utilized as data-collecting devices for the authentic science assessment process. Descriptions and content analysis were used for the evaluation of the gathered data. At the end of the study, it illustrated an innovative application of an authentic science assessment for pre-service science teachers in detail. Such a creative authentic science assessment process based on Cunningham engineering design integrated with Toulmin argument pattern components made pre-service science teachers construct their argument so as to think critically. Also, the illustrated authentic assessment domain would be a multicultural guide for university educators. In the current research, only an authentic science assessment domain was illustrated, this was the limitation of the research, so for further studies, authentic assessment domain constructions could be suggested.

Keywords: authentic science assessment, pre-service science teachers, critical thinking

1. Introduction

In today's modern education systems, the development of critical thinking is considered crucial and necessary. The reason for this might be regarded as the fact that individuals can make fast and correct decisions in preparation for all kinds of conditions that they may encounter in daily life for the vast changes and developments of the age. The main point of making fast and correct decisions depends on individuals' effective use of critical thinking skills. Thus, integrating the development of this skill into science programs is aimed to enable students to think about the daily life problems they encounter and adapt to the age we are in by evaluating these problems from different perspectives. Scientists have made many definitions of critical thinking. Fisher and Scriven (1997), for example, defined critical thinking as "observation and communication, the active interpretation and evaluation of knowledge and argumentation" (p. 21). Fisher (2001) also expressed critical thinking as "the ability to interpret, analyse, and evaluate ideas and arguments" (p.10). As seen from the definitions, in developing critical thinking, the argumentation process should be included in effective decision-making for solving problems. In addition, educators must be competent in developing the necessary thinking skills to support students' decision-making process and assess whether critical thinking is taking place. In this respect, one of the approaches for developing and evaluating critical thinking skills is authentic evaluation environments, which have gained popularity in recent years.

People have been testing and measuring other people for centuries. But what must be the ideal standards for efficient assessment, especially in education? Ten developmentally appropriate practice assessment standards in education in the past decade cleared up as a utility, acceptability, authenticity, equity, sensitivity, convergence, collaboration, congruence, technology, and outcomes. Utility refers to the assessment's usefulness for intervention, acceptability refers to the assessment's social worth, and authenticity refers to natural methods and contexts for assessment. Equity is essential for assessment as it relates to assessment's adaptability for special needs. Sensitivity provides fine measurement gradations, whereas convergence means synthesizing ecological data. Collaboration requires parent and professional teamwork through an assessment process. Congruence allows specially designed, field validated, and/or evidence-based assessment processes. On the other hand, the technology dimension would implicate computer-based observations, recordings, and reporting to the assessment processes. As an outcome, outcome benchmarks would be needed and important (Neisworth & Bagnato, cf., Bagnato, 2007).

The authentic assessment has recently been increasingly used in teaching environments (Maxwell, 1999). Authentic assessment focuses on the practical application of problems encountered in daily life (Fook & Sidhu, 2010). The realistic problem-based teaching domains would engage students with daily life and also would make teachers evaluate students' problem-solving skills by looking at their arguments and prototypes designed for problem-solving (Cunningham, cf. Tozlu et al., 2019). Besides that, using realistic assessment activities in teaching domains would fill the gaps in the teaching curriculums and contribute to increasing students' daily academic and emotional performances (Dunn & Mulvenon, 2009).

The authentic assessment directly examines students' performances on worthy intellectual tasks. Traditional assessment, by contract, relies on indirect items, simplistic substitutes from which it is thought valid inferences can be made about the students' performances at those challenges (Wiggins, 1990, p.1). It is also necessary to clear the advantages of the authentic assessment. The authentic assessment provides critical reflection where the students would reflect on the scientific and social context of his/her work. Authentic assessment enables the student to interrogate their practice and challenge the structures of their work. In the authentic assessment domains, the nature of the relationship between the teacher and students is dialogic, not rhetoric, used as a basis to alter the traditional power hierarchy between them (Rennert-Ariev & College, 2005).

The characteristics of authentic assessment were summarized under three categories: assessment context, students' role, and scoring. The assessment context must be realistic, daily life connected, performance-based, and complex in terms of cognitive aspects (Frey, 2013). For example, In the introductory research methods lesson, a teacher wanted students to prepare a research report for an authentic task to be published in a journal. In the task, students were asked to prepare a report comparing five different products, such as detergent powder, yogurt type, or car insurance, in the form of a scenario in accordance with the structure of authentic nature (Herrington & Herrington, 2006). Also, students must justify ideas or products, be active and formative through the assessment process, and work collaboratively (Frey, 2013). To explain the ideas, an authentic assessment application process could be given as a sweets problem. When someone eats certain sweets, they dissolve in his/her mouth and fizz. They do the same thing in water. What the students have to find out: What makes the difference in how long the sweets last? Is it how hot or cold the water is or how much the sweet is broken up? In this case, students were instructed that they must choose whatever they needed to answer the question. They must make a clear record of their results so that someone else can understand what they have found out (Cumming & Maxwell, 1999). For scoring, the outcomes must be known; in other words, they must be constructed by students, and there must be various assessment ways, such as portfolios. It would be beneficial to be clear that the outcome and the performance must be scored (Frey, 2013).

The authentic assessment must make students gain some higher-order thinking skills. Moreover, by participating in assessment processes, students should establish assessment and evaluation procedures that measure the quality of their learning and work (Luongo-Orlando, 2003). There must be an assessment process to make students think, make critical reasoning, use knowledge from various disciplines, and rapidly set the same questions for much better memorization (Malaysian Examination Syndicate, cf., Mohamed & Lebar, 2017). However, it might be said that authentic assessment practices have some negative features in teaching environments. Those practices involving authentic assessment also require a lot of time and effort for students and teachers. Although work avoids these disadvantages, it is only a trade-off for the quality of learning, motivation, and transfer that result from such activities (Svinicki, 2004).

Effective teachers make students analyse and synthesize knowledge, apply what they have learned to daily life problems, and demonstrate their understanding of material according to determined criteria. They have improved their students' learning and evaluation experiences and taught them how to produce, rather than reproduce, scientific information (Burke, cf., Scott, 2000). The need for making students challenge different assessment contents requires a beneficial and efficient teacher education that would enable pre-service teachers to gain experiences in understanding the effects of contexts. Rather than applying traditional assessment routines, pre-service teachers need to gain experience before in-service in the ability to assess teaching situations, processes, and outcomes and to develop teaching responses that can be effective under different circumstances (Darling-Hammond & Snyder, 2000; Sullivan & McConnell, 2017). Thus, in the current research it was aimed to design an authentic science assessment domain for pre-service science teachers to enhance their higher thinking skills, such as critical thinking skills; being aware that if pre-service science teachers experienced effective authentic assessment domains, then it would be much more probable for them to construct such effective authentic assessment domains for their future students. In the literature, there were no studies integrating Cunningham's (2009) engineering design with Toulmin's (2003) argument pattern components in means of an authentic assessment domain for helping students to think critically so it was thought that the current research would make a contribution to the gap for further multicultural authentic assessment domain design researches.

2. Literature Review

In the literature, there were some studies on authentic assessment processes conducted with pre-service teachers. For example, in their research, Jaelani and Umam (2021) applied authentic assessment to pre-service English teachers. The action research was used in the study, and documentation, observation, questionnaire, and interview were employed as data collection tools. The study results showed that the use of authentic assessment helped the pre-service English teachers understand the nature and use of authentic assessment for their future needs in the profession. Also, Kinay and Bağçeci (2017) studied the effect of the authentic assessment process in scientific research methods lecture on pre-service teachers' beliefs towards learning and participative assessment. Quantitative research was employed in the study, and learning beliefs and beliefs towards participative assessment scales were used as data collection tools. According to their findings, the authentic assessment activities contributed to the pre-service teachers' constructive learning beliefs and reduced their traditional learning beliefs in the experimental group. Moreover, it was shown that authentic assessment activities contributed to the pre-service teachers' beliefs toward participative assessment in the experimental group. In contrast, there was no contribution found in means of learning beliefs or beliefs towards participative evaluation in the control group. Another study investigated the pre-service primary school teachers' emotional features in scientific research methods lectures through authentic assessment processes. Pre-service primary school teachers' diaries were used as data collection tools, and content analysis was used for the gathered data. The study's findings showed that in the first weeks, the pre-service primary school teachers mainly felt negative emotions, and in the last weeks, negative feelings decreased. Instead, positive emotions were replaced (Kinay & Bağçeci, 2015). Kinay and Bağçeci (2014) researched the opinions of pre-service primary school teachers through the authentic assessment process in the scientific research methods lecture. There were 21 participants, semi-structured interviews were done with the participants, and descriptive analysis was employed for the gathered data. At the end of the study, the pre-service primary school teachers stated positive opinions about the authentic assessment process. They also indicated that it was good to evaluate the product of the process and the process itself. Also, Moon et al. (2005) studied an authentic assessment based on learner-centered psychological principles for gifted students whereas Aziz et al. (2020) indicated in their research that there was no clear authentic assessment process flow for teachers and also their research provided some insights on the preparations and the use of authentic assessment as part of teaching and learning process. Thus, being different from the literature, in the current study an innovative authentic science assessment process based on Cunningham's (2009) engineering design process integrated with Toulmin's (2003) argument pattern components was employed for enhancing pre-service science teachers' critical thinking skills, and the whole process was told in detail for the validity and reliability of the whole study. To summarize the novelty of the research was that the authentic assessment process was a step-by-step process in a spiral philosophy on Cunningham's (2009) engineering design process integrated with Toulmin's (2003) argument pattern components.

3. Research Methodology

3.1. Research Design

The current research was a case study design as one of the qualitative research designs. In the case studies, the education programs, the teaching processes, and the people were researched to understand their stories in detail (Stake, 1995). This study, it was investigated pre-service science teachers' critical thinking skills in an authentic science assessment domain in detail. Hence, it was aimed to make an illustration for further multicultural authentic assessment domain designs.

3.2. Participants of the Study

The participants of the study were 34 science teacher candidates studying at a state university located in the eastern part of Turkey. The participants were educated at the university's third science teacher education program class. 27 participants were female, and seven participants were male. Their participation in the study was based on willingness. It was told to the participants by the researcher, who was one of the article's authors, that they could move out of the research process whenever they wanted. After the research, the participants confirmed the research findings for the study's validity and ethics.

3.3. Application Process

In the authentic science assessment process, the participants made six groups, each containing participants between four and seven, to work in collaboration. The authentic science assessment process was conducted for eight weeks, two hours per week. The problem situation was given to the participants from daily life. It was asked the pre-service science teachers to design a vacuum cleaner at least by using two multidisciplinary targets of the science teaching education program for the secondary school students who pre-service science teachers would educate in the future. The vacuum cleaner the participants would design must have some characteristics such as usefulness, economical, long operating time, toughness, and marketableness.

The authentic science assessment process employed Cunningham's (2009) engineering design process. In other words, the authentic science assessment process was constructed on Cunningham's engineering design process, which consisted of asking, imagining, planning, creating, and improving steps. In the 'ask' step, a problem is given to participants. In the 'imagine' step, participants brainstorm and choose the best option among the alternatives. In the 'plan' step, the participants draw their designs and pick the needed materials. In the 'create' step, the participants construct their model and test it; in the 'improve' step, the participants argue how to improve their models. And also, Toulmin's (2003) argument pattern components, claim, data, warrant, backing, and rebuttal, are integrated into these steps through the authentic science assessment process. A claim means a temporary hypothesis from these components, whereas data grounds this claim. On the other hand, a warrant connects the claim and data, and backing justifies the warrant. Rebuttal refers to counter-claims for the claim, warrant, and/or backing.

In the first five weeks, the pre-service science teachers designed their vacuum cleaner prototypes in small groups by arguing their opinions by using Toulmin's (2003) argument pattern components based on Cunningham's (2009) engineering design process steps for solving the given problem in other words for designing a vacuum cleaner according to the shared characteristics. In the sixth week, the groups argued about each other's first vacuum cleaner designs and gave feedback to each other, so in the seventh week, each group revised their initial designs. At last, in the eighth week, each group presented their designs with marketing slogans, and then they evaluated each other's designs as a final step.

3.4. Instruments

Worksheets were utilized as data-collecting devices for the authentic science assessment process.

3.4. Data Analysis Techniques

Descriptions and content analysis were used for analyzing the data. Content analysis means classifying data under some particular themes called codes. These codes are organized under some general themes called categories, and then frequencies and percentages must be calculated per each of the codes and/or categories. The cross-content analysis was made on whether all the data consisted of codes or not (Erickson, 2004). The specifically constructed authentic science assessment process worksheet is given in Figure 1. Three science educators worked independently for the validity and reliability of the worksheet. First, each of the science educators checked the worksheets' content validity. The flow of the application process is summarized in Figure 2.

Then each of the science educators made coding and descriptions of the gathered data independently. The coding and descriptions' consistency were compared, and it was seen that there was no significant difference between the three science educators' coding and description-making.

Authentic Science Assessment Worksheet

Step 1: Ask

Make a design of a vacuum cleaner by using two multidisciplinary targets of the science teaching education program for the secondary school students you would educate in the future. The vacuum cleaner you would design must have specific characteristics such as usefulness, economical, long operating time, toughness, and marketableness.

Which disciplines do you have to use to solve the problem? To solve the same problem, what previous solutions improved in literature and/or in the market? Which knowledge would you need to solve the problem? What are the advantages and/or disadvantages of solving the problem? Which materials could you use to solve the problem? What is your data?

Step 2: Imagine

To solve the problem, brainstorm in small groups. What are your potential claims/temporary hypotheses for solving the problem?

Step 3: Plan

What are your warrants and backings for justifying your claims? What are the materials you have determined for searching your claim? What are your pre-arrangements for your probable design? Continue studying in small groups.

Step 4: Create

Draw and explain your design in detail in small groups. How do you gather data? In other words, how do you test your plan to see whether your claim/temporary hypothesis worked or not? Before trying your design, explain which vacuum cleaner characteristics (usefulness, economic, long operating time, toughness, marketableness) gained ground in design. Explain how your gathered data grounded your claims.

Step 5: Improve

How could you improve your design? Go on studying in small groups. Repeat all the steps one by one from the beginning to now on. What are your rebuttals for the other groups' prototypes useless characteristics?

Step 6: Assessment (extra step)

Evaluate your design based on given vacuum cleaner characteristics, and do the same for other groups' designs. Also, compare your first designs with the improved ones and evaluate the improvements of your designs.

Figure: 1 Authentic science assessment worksheet

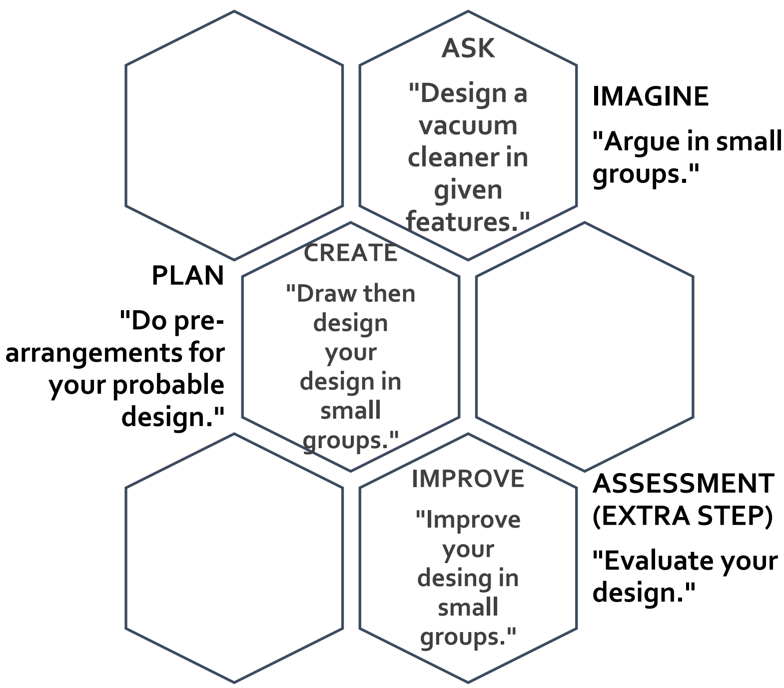

Figure: 2 Authentic science assessment application process flow on the basis of Cunningham (2009) model and Toulmin (2003) argument pattern.

4. Results

4.1. Step 1 Findings: Ask

According to the analysed data gathered from the pre-service science teachers' worksheets, for the first step, it was seen that the whole group could construct content knowledge about how to design the vacuum cleaner according to the previously determined characteristics. They could evaluate the materials they needed to create a vacuum cleaner prototype. Also, they could tell in detail the advantages and disadvantages of each of these materials. Thus, they were successful in grounding their claims on useable data. For instance, group five's descriptive data was "If we want to solve the problem, we need enough content knowledge. Specifically, in physics, we have to know how to construct a circuit, and also, we also have to know and use the basic relation between voltage and power. The advantage of our design would be that our design would work in small pools because, in the three seasons, among the four, there are lots of leaves on pools, so our design would capture these leaves and be portable and useful for everyone. With the remote control, even little kids could use this design. We do not think there are disadvantages to our design, but if water can enter our design's plastic bottles, then the design's engine may not work. But we will improve the mechanism of the design so that this likely problem would be no more threat to the design. We will use plastic bottles, engine, remote control system, fishnet, and skewers." *(Data component of Toulmin's argument pattern)*

Group two's descriptive data was "For designing the electrical vacuum cleaner, it was benefited from the physics and engineering disciplines. There is an electrical engine in electrical vacuum cleaners, and this engine controls a fan. When this fan overturns, it creates a vacuum effect inside the machine. Because of this vacuum effect, air would be pushed inside the machine. While the air moves through the machine, it pulls the little dust into it. The emitted air is pushed out again because of the fan effect each time, but the emitted dust gathers in a bag of the machine. It was needed the knowledge of how an engine works, how the circuit of the machine is built, where the main magnet is, and what the designs of plastic pipe, magnet, and filter are for decomposing and selecting the particles. The advantages of our design are the engine's vacuuming power, the used recycling materials, and the recycling machine itself. The disadvantage of our design is that the machine itself is huge. The materials used are a rubbish bucket, magnet, wheels, tray, filter, engine, cables, switch, double-sided tape, plastic pipes, hot silicon, and styrofoam." *(Data component of Toulmin's argument pattern)*

4.2. Step 2 Findings: Imagine

Each group could construct their claims based on their previously built data. An illustration could be given from group two "Our problem was making an innovative design of an electric vacuum cleaner according to given characteristics. We argued about how to design an answer to the problem by providing the vacuum cleaner's different, useful, economical, and resistant characteristics. The arguing process was a time-consuming process for us. We employed different claims and methods each time, and we modified and revised them each time because of breaking the characteristics. In the end, all of us in the group approved the last claim of us and then we employed this revised claim. We decided to use a hydraulic system in our vacuum cleaner design. There, we designed a different and usable model." *(Claim component of Toulmin's argument pattern).*

On the other hand, group six constructed a basic claim "By the improved vacuum cleaner that could be humped, that is light and has a powerful vacuum system, we could clear up the mess in gardens." *(Claim component of Toulmin's argument pattern)*

4.3. Step 3 Findings: Plan

Each group planned their vacuum cleaner designs according to their previously constructed data and claims using the given characteristics. Therefore, they also were successful in completing their warrants and backings for justifying their claims in planning the prototypes' detailed features after engaging data and claims, as could be seen in Figure 3 *(Warrant and backing component of Toulmin's argument pattern).*

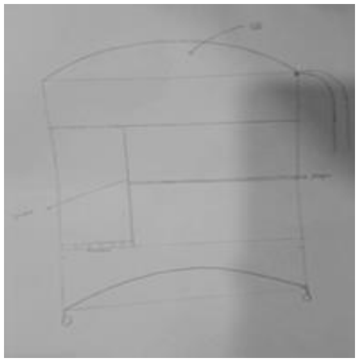

Figure: 3 Group two are planning before constructing their vacuum cleaner prototype

4.4. Step 4 Findings: Create

According to Cunningham's (2006) engineering design process, the pre-service science teachers were able to construct a design based on the given vacuum cleaner characteristics for solving a potential daily-life problem they could come to face in their lives. For example, group two's created design is given in Figure 4. The group explained their plan: "We put a powerful engine at the bottom of a rubbish basket. Then we joined cycling wheels under the rubbish basket using a tray to make the whole design move easily. We put a magnet and filter in it to make our design to be selective for dust and rubbish. There were three plastic pipes in the basket to make our design much more sensitive to dust and rubbish. Our design works with electricity. And also, there was a vacuuming system upwards of the basket." *(Again, data component of Toulmin's argument pattern).*

Figure: 4 Group two's design

Also, group one's design is given in Figure 5. The group explained their design: "The design was constructed with a vacuum system controlled by a machine engine, and a hydraulic system with different-sized heads. Because of these hydraulic arms, there were different-sized functional heads." *(Again, data component of Toulmin's argument pattern).*

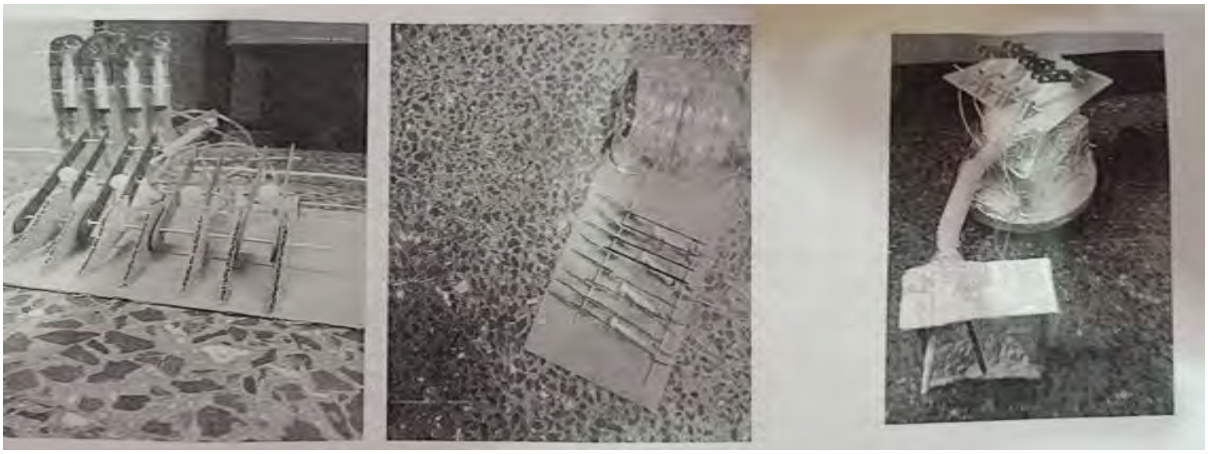

Figure: 5 Group one's design

4.5. Step 5: Improve

The groups successfully finished designing their vacuum cleaner prototypes in the fifth week. In the sixth week, the groups argued about each other's vacuum cleaner designs. They gave feedback to each other, especially regarding useless characteristics of the prototype *(rebuttal component of Toulmin's argument pattern)*, so in the seventh week, each of the groups revised and improved their initial designs.

4.6. Step 6: Assessment (extra step)

During the eighth week, each group presented their designs with marketing slogans; then, they evaluated each other's designs as a final step. The assessments of each of the groups for each other were given in Table 1.

Table: 1 The assessments of each of the groups for each other's vacuum cleaner prototype

Groups' prototypes	**Assessments of the groups**
G1's prototype	*G2's evaluation:* Marketing labour was good. We liked that the prototype was designed on the basic characteristics of environment-friendly, high vacuuming power, low cost, and usable with solar energy and/or batteries. *G3's evaluation:* The portability and usefulness of the vacuum cleaner prototype were not enough. *G4's evaluation:* There could have been holding places in the prototype, but there was no. Its movement was less, so vacuuming the dust capacity would not be efficient. The plastic pipes could be in a bit downwards and could be a little small. *G5's evaluation:* The cost was average. But it could be a bit noiseless. The visuality of the prototype was good. But only vacuuming small particles characteristic could be a disadvantage because its usage would decrease. There was no switch, which was a disadvantage because it only worked when it was plugged in. *G6's evaluation:* We thought its cost was much. We could not see that the prototype was

	successful at vacuuming the dust. The prototype did not work, even if it did not move. Only its engine was working, and nothing more...
G2's prototype	*G1's evaluation:* They told us about their vacuum cleaner prototype that there was a solar panel, but there was not any. *G3's evaluation:* There was no movement, only cleaning its surroundings. It was making so much noise that it could be a bit noiseless. *G4's evaluation:* The prototype's noise was so bothersome, and its cost was average. But the energy source for making the vacuum cleaner work was a problem because, outdoors, finding an energy source was not so probable. It was made of recycling materials that we approved so much. *G5's evaluation:* The vacuum cleaner prototype's potential area for cleaning was small. There was no remote-control system for controlling the vacuum cleaner's arms, and the arms worked independently. The prototype could have been much more controllable. *G6's evaluation:* The recycling material usage for demonstrating the prototype was meaningful. But the prototype could have been designed as moveable.
G3's prototype	*G1's evaluation:* This prototype design was very laborious, and it was not practical. And also, it was not working, and it was dysfunctional. There were unnecessary and dysfunctional materials in the design, and its working mechanism was not suitable and it was not resistant. *G2's evaluation:* The prototype's cost was low, but we did not think that it was profitable and productive in means of use. Its outlook fitted well, and its mechanism was thought well, but it was not working because the materials connections in a circuit were not established well. *G4's evaluation:* The prototype was not productive. It could not vacuum so much dust, and its engine power was insufficient. *G5's evaluation:* The portability of the constructed vacuum cleaner was not practical. The hydraulic system must be integrated into a pipe so its usage can be much more efficient. And also, if a wheeling system was integrated into the prototype, the transporting could be much easier. *G6's evaluation:* The prototype could be helpful in means of cost.
G4's prototype	*G1's evaluation:* The prototype's rechargeable battery at once and its long life were economical in terms of efficient energy usage. But the rubbish bag was small, so this would cause frequent cleaning. On the other hand, if it was made much bigger, then the prototype could be no more useful in means of the given characteristic of the prototype. *G3's evaluation:* The fans of the prototype were bought off the self and modified to the prototype. Its movement was limited, and sometimes it stopped moving. The prototype no more vacuumed the dust. Instead, they were under the prototype. The whole system was not working at full performance. All these were the disadvantages. And the design's effort and labour were not so much. *G5's evaluation:* The prototype's engine was not so powerful; it was only capable of vacuuming so small particles. Its cost was not so much. The take-off and putting again on the apparatus of the machine were difficult, and its working area was limited. *G6's evaluation:* The prototype was introduced as useful for disabled people. But it could only help walk disabled people, not others.
G5's prototype	*G1's evaluation:* The idea of the design was so innovative. The cost of the prototype was low. But its movement area was limited. *G2's evaluation:* Nevertheless, the prototype usage was average and could still be used. *G3's evaluation:* Water contact could make the design out of use. And also, the filter could easily get out of use. *G4's evaluation:* The prototype's usable area was limited, and it was economical, but still,

	it needed so badly to improve. *G6's evaluation:* The engine system of the prototype was visible; instead, it could be put into a closed system. The design was economic.
G6's prototype	*G1's evaluation:* The prototype's engine was powerful and also usable. But it could cause backache. It was suitable for the prototype because it could be used with different energy sources. *G2's evaluation:* After a time, the prototype would cause a backache, it was a handicap, and there was no action taken. The cable of the prototype was a handicap for the usage too. *G3's evaluation:* The portability of the prototype in the back of a human was a big problem because after a while the felt heaviness would increase relatively. And also, the heaviness of the prototype would increase in reality because of the rubbish too. The prototype had a cable, but it was a handicap for its usage in big places such as gardens. *G4's evaluation:* The design's outlook fitted well. Its cost was low. The idea of the design was innovative. But its only way of working based on an electrical cable system was a problem because its usage outdoors would be the problem. *G5's evaluation:* There could be a reel system in the prototype. Its size was too big. After a time, carrying the design back, in other words, humping would be a problem because it would cause individuals to be exhausted. Designing labour and effort were less.

As seen in Table 1, all the groups successfully constructed the advantages and disadvantages of each of the prototypes. In other words, the pre-service science teachers were good at making successful self-evaluations of previously determined characteristics for vacuum cleaner prototype design.

Ask, imagine, plan, create, improve steps and assessment step as an extra step based on Cunningham's (2009) engineering design process integrated with Toulmin's (2003) argument pattern components throughout the authentic assessment process for helping pre-service science teachers to think critically were described in detail for giving a chance to university educators for further researches in means of multicultural applications.

5. Discussion

The current research illustrated an innovative application of an authentic science assessment for pre-service science teachers to enhance their critical thinking skills. The authentic science assessment process was constructed on Cunningham's (2009) engineering design process, which consisted of the ask, imagine, plan, create, and improve steps. And also, Toulmin's (2003) argument pattern components, claim, data, warrant, backing, and rebuttal, integrated into these steps through the authentic science assessment process for enhancing pre-service science teachers' critical thinking skills. In the current research, it was thought that if pre-service science teachers experienced effective authentic assessment domains, then it would be much more probable for them to construct such effective authentic assessment domains for their future students. Also in the current research, it was thought that a detailed description of the authentic assessment domain for pre-service science teachers' education to make them think critically so to be able to design alike domains for their future students, would give an illustration for multicultural authentic assessments too. In other words, it was also aimed to give an illustration, to make a guide for university educators from all around the world who would want to construct different authentic assessment domains for their pre-service science teachers to help them be critical thinkers.

The study made pre-service science teachers design their own prototypes throughout an authentic assessment domain by integrating Cunningham's (2009) engineering design with Toulmin's (2003) argument pattern components, to help them think critically. In other words, it could be said the pre-service science teachers' engineering design ability and their critical thinking ability were enhanced. Just like Zhong et al. (2023) utilized robotic applications in education, as an output their students' engineering design ability and cognitive load were enhanced.

In the study, through the authentic science assessment process' Cunningham engineering model steps, it was made pre-service science teachers construct scientifically valid arguments in groups via collaboration step by step. According to the Cambridge International Thinking Skills Syllabus 2020-2022 vision, if individuals could build scientifically valid arguments, then it would mean they could think critically, which was a higher-metacognitive thinking skill. Thus, based on the Cambridge International Thinking Skills Syllabus 2020-2022 vision, it could be said that this illustrated authentic science assessment process helped pre-service science teachers enhance their critical thinking skills. Also, according to literature, the authentic assessment must make students gain some higher-order thinking skills such as critical reasoning, use of learning from various disciplines, set the same questions rapidly for much better memorizing (Malaysian Examination Syndicate, cf., Mohamed & Lebar, 2017), and in the current study, the pre-service science teachers questioned all the Cunningham engineering model steps so to make critical reasoning than to construct arguments. Also, through Cunningham's (2009) engineering model, the pre-service science teachers had the chance to experience a problem, brainstorm about it, plan the prototype, create a prototype, and improve the prototype as engineering skills, just like as in Arık and Topcu's (2023) study, through identification of the problem, generation of design ideas, and creation and improvement of the final design artifact steps, middle school students' engineering design process skills were studied.

The characteristics of authentic assessment were highlighted in the literature that assessment context must be realistic, daily life connected, performance-based, and complex in terms of cognitive aspects (Frey, 2013). Hence, in the current study, the pre-service science teachers were given a real, daily life-connected, performance-based, and complex problem: designing a vacuum cleaner with the characteristics of usefulness, economical, long operating time, toughness, and marketableness. According to the literature, the students must justify ideas or products, they must be active and formative through the assessment process, and also, they must work collaboratively (Frey, 2013), which on the whole were illustrated through all steps in this current authentic science assessment process step by step based on Cunningham engineering design and told in detail through the article. For scoring, there must be various assessment ways, such as portfolios. It would be beneficial to be clear that the outcome and the performance must be scored (Frey, 2013). Like in the current study, the whole process was evaluated in detail based on Cunningham's engineering design. The benefits of authentic assessment could be summarized as a shift for students from isolation of society to being a member of the society (McArthur, 2023), preparation of the students for future workforce (Nyanjom et al., 2023), improvement of students IT levels (Hidayah et al., 2023), higher cognitive skills (Ifelebuegu, 2023), and self-efficacy (Ismail et al., 2023). Thus, the current research underlying once more the critical thinking output as a higher cognitive skill would be valuable.

In literature, Purzer et al. (2022) underlined the importance of engineering education for students from all levels. In the current research, it was modeled a detailed engineering modeling process in detail, but only one engineering modeling process as authentic assessment was studied in detail, it was a limitation of the study, and for further studies, different engineering modelings as authentic assessment domains could be offered. Also, in this study, only the university level was studied, education for students from all levels, from kindergarten through 12th grade, could be offered too. For example, in literature elementary students' engineering design activities showed that the students enjoyed the process, above all (Yang & Chittoori, 2022). In the current study authentic assessment domain was illustrated through engineering design processes for improving the students' critical thinking, but for improving students' engineering design skills, alternative teaching domain designs and assessment of them as project-based learning in the laboratory just as in literature (Gomez-del Rio & Rodriguez, 2022), could be offered also. Finally, the focus of this research was on prospective science teachers, who are the teachers of the future. However, in the century we live in, there is a need for people from different cultures to live together and to seek solutions together for real-life problems encountered, both due to the dizzying development of technology, the increase in international communication, and the experience of migration due to reasons such as natural disasters, economic and political crises, epidemics. This is also seen in educational settings. In recent years, the situation of students from different cultures receiving education together in classrooms has increased considerably. In this respect, students from different cultures will come up with solutions for the problem situations they encounter in daily life together in the future. In many scientific organizations, scientists from different cultures conduct scientific studies together. Therefore, evaluating the existing situation from all angles is recommended by designing educational environments where individuals from different cultures can use their diverse skills and work together in educational settings.

6. Conclusion

It was seen in the current study that an innovative authentic science assessment process based on Cunningham's (2009) engineering design process integrated with Toulmin's (2003) argument pattern components for enhancing pre-service science teachers' critical thinking skills, was an efficient one, in other words, the aim of the study was ensured, so, for further studies different innovative and detailed authentic assessment processes could be offered for enhancing pre-service teachers' high metacognitive thinking skills.

REFERENCES

Arık, M., & Topçu, M.S. (2023). Development and implementation of engineering-based aircraft unit: Middle school students' engineering design process skills. *International Journal of Technology and Design Education.* https://doi.org/10.1007/s10798-023-09829-7.

Aziz, M.N.A., Yusoff, N.M., & Yaakob, M.F.M. (2020). Challenges in using authentic assessment in 21st-century ELS classrooms. *International Journal of Evaluation and Research in Education*, *9*(3), 759-768.

Bagnato, S.J. (2007). *Authentic assessment for early childhood intervention.* The Guilford Press.

Cambridge International Thinking Skills Syllabus 9694 (2020-2022). *https://www.cambridgeinternational.org.*

Cumming, J.J., & Maxwell, G.S. (1999). Contextualizing authentic assessment. *Assessment in Education: Principles, Policies and Practices*, *6*(2), 177-194.

Cunningham, C.M. (2009). Engineering is elementary. *The Bridge*, *30*(3), 11-17.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, *16*, 523-545.

Erickson, E. (2004). Demystifying data construction and analysis. *Anthropology and Education*, *35*(4), 486-493.

Dunn, K.E., & Mulvenon, S.W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical, Assessment, Research, and Evaluation*, *14*, 1-11.

Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment.* University of East Anglia, Centre for Research in Critical Thinking.

Fisher, A. (2001). *Critical thinking: an introduction. What is critical thinking and how to improve it?* Cambridge: Cambridge University Press.

Fook, C.Y., & Sidhu, G.K. (2010). Authentic assessment and pedagogical strategies in higher education. *Journal of Social Sciences, 6*(2), 153-161.

Frey, B.B. (2013). *Modern classroom assessment*. SAGE Publications.

Gomez-del Rio, T., & Rodriguez, J. (2022). Design and assessment of a project-based learning in a laboratory for integrating knowledge and improving engineering design skills. *Education for Chemical Engineers, 40*, 17-28.

Herrington, J., & Herrington, A. (10-12 July 2006). *Authentic conditions for authentic assessment: Aligning task and assessment.* 29th Herdsa Annual Conference, Perth, Western Australia.

Hidayah, F.U., Linuwih, S., & Subali, B. (2023). The effectiveness of "Sipentik" as an authentic assessment application and for increasing student IT literacy. *Journal of Innovative Science Education*, *12*(1), 107-116.

Ifelebuegu, A.O. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning & Teaching*, *6*(2), 1-8.

Ismail, S.M., Nikpoo, I., & Prasad, K.D.V. (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. *Language Testing in Asia*, *13*, 1-20.

Jaelani, A., & Umam, A. (2021). Preparing EFL pre-service teachers for curriculum 2013 through authentic materials and assessment integration. *Journal of English Educators Society*, *6*(1), 171-177.

Kinay, İ., & Bağçeci B. (2014). Sınıf öğretmeni adaylarının otantik değerlendirme sürecine yönelik görüşlerinin incelenmesi. *Elektronik Eğitim Bilimleri Dergisi*, *3*(6), 39-50.

Kinay, İ., & Bağçeci B. (2015). Otantik değerlendirme sürecine katılan öğretmen adaylarının duyuşsal özelliklerinin incelenmesi. *Eğitim ve İnsani Bilimler Dergisi: Teori ve Uygulama*, *6*(12), 67-80.

Kinay, İ., & Bağçeci B. (2017). Otantik değerlendirme yaklaşımının öğretmen adaylarının öğrenmeye ve katılımcı değerlendirmeye yönelik inançlarına etkisinin incelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, *52*, 16-32.

Luongo-Orlando, K. (2003). *Authentic assessment: designing performance-based tasks*. Pembroke Publishers.

Maxwell, G.S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policies and Practices, 6*(2), 177-194.

McArthur, J. (2023). Rethinking authentic assessment: work, well-being, and society. *Higher Education*, *85*, 85-101.

Mohamed, R., & Lebar, O. (2017). Authentic assessment in assessing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences*, *7*(2), 466-476.

Moon, T.R., Brighton, C.M., Callahan, C.M., & Robinson, A. (2005). Development of authentic assessments for the middle school classrooms. *The Journal of Secondary Gifted Education*, *16*(2/3), 119-133.

Nyanjom, J., Goh, E., Yang, E.C.L. (2023). Integrating authentic assessment tasks in work integrated learning hospitality internships. *Journal of Vocational Education & Training, 75*(2), 300-322.

Purzer, Ş., Quintana-Cifuentes, J., & Menekse, M. (2022). The honeycomb of engineering framework: Philosophy of engineering guiding precollege engineering education. *Journal of Engineering Education, 111*, 19-39.

Rennert-Ariev, P., & College, L. (2005). A theoretical model for the authentic assessment of teaching. *Practical Assessment, Research and Evaluation, 10*(2), 1-11.

Scott, J. (2000). Authentic assessment tools. In R.L. Custer (Ed.), J.W. Schell, B. McAlister, J. Scott, & M. Hoepfl. *Using authentic assessment in vocational education.* Information Series No. 381 (pp. 40-55). Eric Document Reproduction Service No. Ed 440 293.

Stake, R.E. (1995). *The art of case study research.* SAGE Publications.

Sullivan, D.F., & McConnell, K.D. (2017) Big progress in authentic assessment, but by itself not enough. *Change: The Magazine of Higher Learning*, *49*(1), 14-25.

Svinicki, M.D. (2004). Authentic assessment: Testing in reality. In M.V. Achacoso & M.D. Svinicki (Eds.), Alternative Strategies for Evaluating Student Learning. New Directions for Teaching and Learning, 100, 23-29. Jossey-Bass.

Toulmin, S. (2003). *The uses of argument.* Cambridge University.

Tozlu, İ., Gülseven, E., & Tüysüz, M. (2019). FeTeMM eğitimine yönelik etkinlik uygulaması: Kuvvet ve enerji örneği. *YYÜ Eğitim Fakültesi Dergisi, 16*(1), 869-896.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*(2), 1-3.

Yang, D., & Chittoori, B. (2022). Technology-supported engineering design and problem solving for elementary students. *International Journal of Technology in Education and Science, 6*(4), 524-542.

Zhong, B., Zheng, J., & Zhan, Z. (2023). An exploration of combining virtual and physical robots in robotic education. *Interactive Learning Environment, 31*(1), 370-382.

Acknowledgements

This study was supported by Van Yüzüncü Yıl University Scientific Studies Project Coordination Centre-Türkiye with the ID number of SAP-2019-8542.