

## Designing and Developing a Test Specification Table for 6th Grade Science Using Software Architecture, Combining “Marzano, Kendall”, and “Anderson et al” Approaches

Roghaye Asadi\* 

*Corresponding Author*, Department of Educational Measurement, Allameh Tabataba'i University, Tehran, Iran. E-mail: r\_asadi@atu.ac.ir

Jalil Younesi 

Associate Professor, Department of Educational Measurement, Allameh Tabataba'i University, Tehran, Iran. E-mail: younesi@atu.ac.ir

Mohammad Asgari 

Associate Professor, Department of Educational Measurement, Allameh Tabataba'i University, Tehran, Iran. E-mail: drmasgari@atu.ac.ir

Enayatollah Zamanpour 

Assistant Professor, Department of Educational Measurement, Allameh Tabataba'i University, Tehran, Iran. E-mail: zamanpour@atu.ac.ir

### Abstract

The goal of this research was to create and develop a table of test specifications based on the approaches of "Marzano and Kendall" and "Anderson et al." for the sixth-grade basic science curriculum, using a developmental method. The study population encompassed all web-based testing tools and experts in the fields of assessment and measurement, software, and science. This article outlines two phases of the research and development process, taking an architectural perspective into account. The first phase involved a thorough analysis of needs and background, while the second phase encompassed the design, architecture, and formative evaluation. In the initial phase, we examined research literature and software related to question design. In the subsequent phase, two software prototypes were developed. The first prototype established the initial software framework based on a combined approach, employing artificial intelligence chatGPT-4 to generate 406 behavioral verbs and 5137 behavioral objectives, whose validity was confirmed by assessment and measurement experts. In the second prototype, the software architecture was designed as a web application using the PHP language for the backend, the Laravel framework, JavaScript language, and the React framework for the front end, along with a MySQL database and the REST protocol for server communication. This phase also addressed and resolved the initial sample's issues and bugs. Furthermore, the input and feedback from other experts were incorporated into the final design.

**Keywords:** Combined table of test specification, Anderson et al., Marzano and Kendal, Software architecture Developmental research

**Cite this Article:** Asadi, R., Younesi, J., Asgari, M., & Zamanpour, E. (2024). Designing and Developing a Test Specification Table for 6th Grade Science Using Software Architecture, Combining “Marzano, Kendall”, and “Anderson et al” Approaches. *Educational Psychology*, 20(71), 7-38. <https://doi.org/10.22054/jep.2024.76280.3933>



## 1. Introduction

This paper aims to examine the role and benefits of a table of test specifications in designing questions for high levels of learning. A table of test specifications is a matrix that shows how questions are aligned with learning objectives and their weights in the curriculum. The paper reviews various classifications of learning objectives, such as Bloom's taxonomy and its revisions by Anderson et al. (2001) and Marzano and Kendall (2007). These classifications are hierarchical and focus on different dimensions and levels of cognitive skills. Using a table of test specifications can ensure the validity and quality of the assessment and align it with educational objectives. The paper also discusses the use of different software and websites for designing online questions and the lack of attention to the cognitive levels and learning objectives in these tools. The paper proposes a new software that can design questions according to the behavioral verbs and definitions based on Bloom's revised taxonomy and Marzano and Kendall's classification. The software can have features such as providing feedback and preventing output if the question is not suitable for measuring the desired level. The software is intended for designing test questions for science courses, considering their importance and role in shaping higher levels of education.

### Research Question(s)

- What are the features of a valid and practical supportive tool that has the potential effectiveness on teachers' performance for empowering them in designing high quality questions in assessing at a high cognitive level?
- Can the software-based educational assessment tool solve the challenges of designing questions involving higher cognitive levels?
- Can the software-based educational assessment tool solve the challenges of enhancing teachers' abilities in designing high-quality questions?

This includes:

- Who should be the target audience?
- How should this tool be accessible to the audience?
- What aspects should the software support for the assessment development?

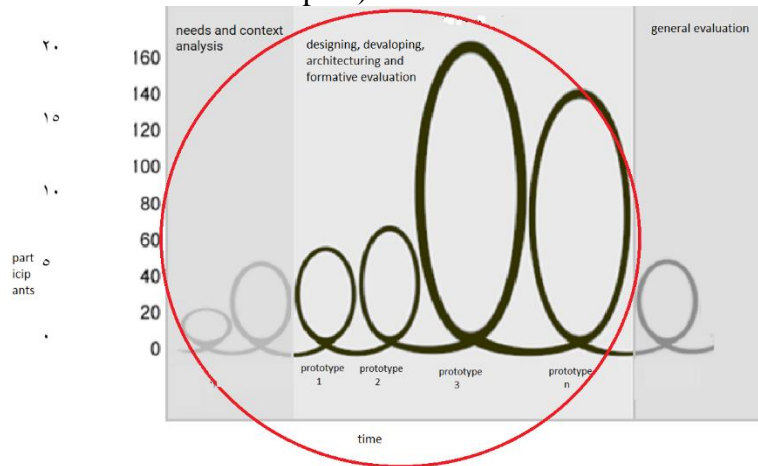
## **2. Literature Review**

The classification of educational objectives by Anderson et al. (2001) is based on Bloom's classification but with some changes. The classification has two dimensions: knowledge and cognitive process. Knowledge includes four types: factual, conceptual, procedural, and metacognitive. The cognitive process includes six types: remembering, understanding, applying, analyzing, evaluating, and creating. Marzano and Kendall's (2007) classification of learning tasks is based on the level of awareness and mental control. It describes three mental systems: self, metacognitive, and cognitive, and how they interact with knowledge to perform a new task. An example is a student who has to decide whether to pay attention to science class or social activity. Each system plays a role in the decision and the action. The classification has two dimensions: mental processes and knowledge. Mental processes have six levels, from retrieval to self-system, and knowledge has three domains: information, mental processes, and psycho-motor processes. Based on these two classifications, the research designed and developed a combined classification that can describe the complexity of learning tasks more accurately.

## **3. Methodology**

This paper used a developmental method for designing and developing a software tool that helps teachers create science questions for elementary students based on a table of test specifications. The method consists of three main stages: needs and context analysis, design and development, and evaluation. Each stage has multiple cycles of activities that involve different groups of participants, such as experts in assessment and measurement, elementary education specialists, teachers, and software specialists. The paper is based on a doctoral dissertation on the same topic, so it focuses on the first two stages, which are needs and context analysis and design, architecture, and formative evaluation. The needs and context analysis stage aims to gain knowledge about the goals, user groups, and domains of the software tool. This stage includes reviewing relevant literature, interviewing experts and specialists, and visiting various question design software tools. The findings of this stage are used to design the initial model of the software tool and to get feedback from experts and teachers. The design, architecture, and formative evaluation stage uses the prototyping process to develop the software tool iteratively. The

prototyping process is a way to involve the target group and experts in the design and development activities. The figure below shows the three main stages of the study (analysis, design, and evaluation) and the scope of the cycles within the main stages (in terms of the participants involved and the time spent).



#### 4. Results

The first stage:

1. The first stage includes needs and context analysis of software development.
2. The design, architecture and formative evaluation

##### Prototype 1:

In the continuation of the software development process, second stage of the software development process involves designing the patterns and cycles that make up the software. The first prototype of the software was based on modern approaches to learning objectives, and a combined classification was created by integrating two existing classifications from Anderson et al. (2001) and Marzano and Kendall (2007). So the cognitive, metacognitive and self-systems were considered in the design of the software. In the continuation for defining behavioral objectives for a combined classification of Anderson et al. and Marzano and Kendall, using an artificial intelligence chatGPT-4 feature. The method involves the following steps:

- Collecting 127 behavioral verbs from theoretical sources for each class

- Using chatGPT-4 to produce English equivalents of each verb, increasing the number of verbs to 406
- Using chatGPT-4 to produce definitions, similarities, differences, examples, and applications of each verb in the field of education and learning
- Using chatGPT-4 to produce the steps of performing each verb in detail and step-by-step
- Using chatGPT-4 to produce an example of a behavioral objective for each verb for each class
- Using chatGPT-4 to produce an example as an exercise for each verb

An example for the verb (gives an atypical example) is as follows:

- Behavioral objective: The student will be able to give an atypical example according to the following instructions about vocabulary in science class. The steps of giving an atypical example are as follows:
  - 1- Define the topic: First, you need to define the topic that you are going to provide an atypical example for. This can be anything from a specific group to a general concept.
  - 2- Understand common assumptions: Then you need to have enough understanding of the assumptions or common ideas about this topic. This can be obtained by research, study, or observation.
  - 3- Choose an atypical example: In this step, you need to choose an example that is not usually used in describing this topic or group. This example should be appropriate and valid, but at the same time challenge common ideas.
  - 4- Explain the example: Next, you need to explain your example in detail and completely, showing how it goes beyond common assumptions.
  - 5- Emphasize why it is atypical: Finally, you need to emphasize why your example is atypical and how it can change our understanding of the topic. This part can include a critical review of common assumptions.

The process of producing behavioral objectives for a combined classification of Marzano and Kendall's and Anderson et al. using an artificial intelligence chatGPT-4 feature involves the following steps:

- Producing 5137 behavioral objectives for different classes using chatGPT-4

- Preparing a checklist to verify the validity of the verb classification and the quality of the behavioral objectives
- Designing a software algorithm to implement the method

Due to the extent of the work done, it is not possible to show all its views in this space; but for showing part of the work, the below image is shown. Software schema was designed with teachers' feedback and some of their ideas were applied in the prototype.

### **Software architecture:**

The software architecture of a test specification was designed based on a web application for online access through web browsers on various devices and operating systems. Web apps eliminate the need for installation and provide dynamic interactions, such as forms, enhancing user engagement and promoting business growth. They enable data storage on online servers and centralized updates for users to access the latest version. The web app consists of two main components: the client or front-end, responsible for user interaction and visible elements, and the server or back-end, managing logic and processing. Programming languages like JavaScript, PHP, and frameworks such as React and Laravel are used for implementation. Communication between the client and server is facilitated through REST APIs. The combination of PHP and MySQL for data storage offers robust tools for web development, meeting the requirements of modern web applications.

### **The prototype (2)**

#### **The first cycle**

The evaluation performed on the first output of the designed software had a set of product specifications that were used for further development. The initial prototype helps users to design questions and then open the output in a word processor and print them. Some suggestions and corrections were given in the steps to reach the final specification table to evaluate the initial prototype, and meetings were held with software experts to fix bugs and misunderstandings in the initial prototype.



Complex Combination Procedures	Simple Combination Procedures	Foundational Procedures	Macroprocedures	Tactics	Behavioral objective: The student will be able to "choose" an option according to the following instruction, in the exercise related to the vocabulary of the science lesson. To choose something, you can follow these steps:	چون	
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	1- <b>Identify the options:</b> First, you need to identify the different options available. For example, if you want to buy a book, you can go to the library, bookstore or online sites.	behavioral objective	to choose
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	2- <b>Collect information:</b> For each option, collect the necessary information. For example, for a book, you can check the reviews, summary, author and genre of the book.	behavioral objective	to elect
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	3- <b>Compare and analyze:</b> Compare the options with each other and analyze them. For example, for a book, you can compare the price, content, writing style and author's style.	behavioral objective	to opt for
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	4- <b>Evaluate the pros and cons:</b> For each option, examine the pros and cons. It is better to make a list of the pros and cons of each option and compare them with each other.	behavioral objective	to pick
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	5- <b>Consider personal preferences:</b> Consider your personal preferences. You may have specific criteria that are important to you and that you should consider in your choice. For example, if you are interested in a particular genre of books, this may play a role in your decision.	behavioral objective	to select
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	6- <b>Decide:</b> After the above steps, you have to decide. Based on your analysis, the pros and cons of each option and your personal preferences, choose an option.	behavioral objective	to mention
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	7- <b>Act:</b> After deciding, proceed with the choose operation. For example, if you have chosen a book, buy it or borrow it from the library.	behavioral objective	to acknowledge
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	By following these steps, you can choose an option more accurately and confidently.	behavioral objective	to discern
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Example:	behavioral objective	to detect
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	The student will be able to use the instruction mentioned in the exercise "choose vocabulary" to choose an option.	behavioral objective	to perceive
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 1: Identify the options Example: Finding different options through dictionary, online sources or language books.	behavioral objective	to spot
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 2: Collect information Example: Collecting information related to each word, including translation, meaning, examples of usage and pronunciation.	behavioral objective	to identify
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 3: Compare and analyze Example: Comparing different meanings and concepts of each word and analyzing its applications in sentences and texts.	behavioral objective	to recognize
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 4: Evaluate the pros and cons Example: Examining the pros and cons of each word, such as the number of applications, complexity and usability in daily conversations.	behavioral objective	to realize
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 5: Consider personal preferences Example: Considering personal preferences such as using in formal or informal situations, or using in public or specialized conversations.	behavioral objective	to define
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 6: Decide Example: Choosing words based on the analysis done, the pros and cons of each word and personal preferences.	behavioral objective	to recur
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective	Step 7: Act Example: Using the chosen words in conversations, writing or language exercises.	behavioral objective	to redo
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	to replicate
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	to repeat
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	to enumerate
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	to list
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	complete
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	کامل می کنی
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	fill in
behavioral objective	behavioral objective	behavioral objective	behavioral objective	behavioral objective		behavioral objective	کامل می کنی

## The second cycle

During the screening of processes, a main model of the software was prepared. Meetings were held online and in person between the developer and other experts in different fields mentioned to review the designed structure. This happens at the beginning of the second cycle and the ideas obtained after registration and review were considered in the design of the software.

## 5. Discussion

To answer the first question, an extensive literature review and assessment of question design software were done. The outcome was the development of a tool that harnesses test specification capabilities, and helps in assessing higher cognitive levels and content representation. Various studies support the use of technology in education, highlighting its role in test design and enhancing students' learning experiences (Alifard and Kamyab, 2013; Sohrabloo, 2017, 2018; Newman et al., 2013; McLaughlin et al., 2005).

To answer the second question, the study explores designing educational assessments using software with a focus on higher cognitive levels. A prototype-based iterative approach was used, integrating Anderson et al. and Marzano and Kendall's combined classification. Artificial intelligence aided in refining behavioral objectives, resulting in 5137 behavioral objectives. Teachers' feedback guided software enhancements, aligning with prior research

demonstrating effective test organization and assessment of deep subject knowledge (Patel et al., 2016).

Question 3 of this research explores the potential of a software-based educational assessment tool to improve teachers' question design skills. The study, informed by Patel et al (2016), revealed that the tool, Test Blueprint, effectively assigns topic weights, aligns questions with learning objectives, and reduces examiner variability. To enhance user accessibility, the software was designed as a web app, offering easy updates and cross-platform compatibility. It also allows customization of test content and offers diverse question formats, enriching the assessment development process. These enhancements signify a concerted effort to advance assessment practices in education.

## **6. Conclusion**

Developing software with appropriate quality based on user needs and scientific principles is one of the important tasks in software development. The features, architecture, and design of the software, as well as the methods to eliminate the weaknesses and challenges that may affect the software quality, must be considered. In this paper, we used artificial intelligence and expert opinions to achieve the best solution for the software. Therefore, we aimed to design a type of software with acceptable quality based on scientific and practical principles by identifying user needs with a suitable architecture in order to meet them.



## طراحی و توسعه نرم‌افزار جدول مشخصات آزمون تلفیقی (رویکرد "مارزانو و کندال" و "اندرسون و همکاران") برای درس علوم پایه ششم ابتدایی با چشم‌انداز معماری نرم‌افزار

رقیه اسدی\*

نویسنده مسئول، دانشجوی دکتری رشته سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: r\_asadi@atu.ac.ir

جلیل یونسی

دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: younesi@atu.ac.ir

محمد عسگری

دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: drmasgari@atu.ac.ir

عنایت اله زمانپور

استادیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: zamanpour@atu.ac.ir

### چکیده

این پژوهش با هدف طراحی و توسعه نرم‌افزار جدول مشخصات آزمون تلفیقی (رویکرد مارزانو و کندال و اندرسون و همکاران) برای درس علوم پایه ششم ابتدایی از رویکرد توسعه‌ای استفاده کرد. جامعه شامل ابزارهای آزمون‌سازی تحت وب و متخصصان سنجش و اندازه‌گیری، نرم‌افزار و علوم بود. دو فاز از سه فاز رویکرد توسعه‌ای (با چشم‌انداز معماری) در این مقاله مورد نظر بوده است: فاز اول شامل تحلیل نیازها و زمینه و فاز دوم شامل طراحی، معماری و ارزیابی تکوینی بود. در فاز اول، ادبیات پژوهش و نرم‌افزارهای مرتبط با طراحی سؤال مورد بررسی قرار گرفت. در فاز دوم، دو پروتوتایپ برای نرم‌افزار طراحی شد. در پروتوتایپ اول، شمای اولیه نرم‌افزار بر اساس رویکرد تلفیقی طراحی شد و با استفاده از هوش مصنوعی chatGPT-4، ۴۰۶ فعل رفتاری و ۵۱۳۷ هدف رفتاری تولید شد که روایی آن‌ها توسط متخصصان سنجش تأیید شد. در پروتوتایپ دوم، معماری نرم‌افزار بر اساس وب اپ با استفاده از زبان PHP، و فریم‌ورک لاراول برای بک‌اند، و زبان javascript، و فریم‌ورک react برای فرانت‌اند، و پایگاه داده MySQL و پروتکل REST جهت ارتباط با سرور طراحی و باگ‌ها و اشکالات نمونه اولیه برطرف شد. همچنین نظرات سایر متخصصان در طراحی نهایی لحاظ گردید.

کلیدواژه‌ها: تحقیق توسعه‌ای، جدول مشخصات آزمون تلفیقی، رویکرد اندرسون و همکاران، رویکرد مارزانو و کندال، معماری نرم‌افزار

استناد به این مقاله: اسدی، رقیه، یونسی، جلیل، عسگری، محمد، و زمانپور، عنایت اله. (۱۴۰۳). طراحی و توسعه نرم‌افزار جدول مشخصات آزمون تلفیقی (رویکرد "مارزانو و کندال" و "اندرسون و همکاران") برای درس علوم پایه ششم ابتدایی با چشم‌انداز معماری نرم‌افزار. فصلنامه روان‌شناسی تربیتی، ۲۰(۷۱)، ۳۸-۷. <https://doi.org/10.22054/jep.2024.76280.3933>

## مقدمه

یکی از اهداف اولیه ارزیابی عملکرد دانش‌آموزان، به دست آوردن نتایج درست و واقعی است که نشان‌دهنده دست‌یابی به شایستگی‌های در نظر گرفته‌شده در هدف‌های آموزشی موردنظر است (McLaughlin et al., 2005). منظور از سنجش کلاسی نیز تمام فرآیندهایی است که برای تصمیم‌گیری درباره یادگیری دانش‌آموزان صورت می‌پذیرد (Nitko, 2004)؛ (McMillan, 2007, cited by Eggen & Kauchak, 2010) Marso and Pigge (1993) معتقدند ارزیابی‌های آزمون‌های کلاسی در طول دهه‌های مختلف نشان داده است که بیشتر آزمون‌های معلم‌ساخته تنها به یادآوری اطلاعات مربوط است (cited by Brookhart, 2010)؛ با این حال وقتی از ایشان پرسیده شد که نظر آن‌ها در خصوص آزمون‌های معلم‌ساخته چیست؟ تصور ایشان بر آن بود که سطوح شناختی به کار بستن، منطق، دلیل‌آوری و سطوح بالای تفکر را تا حدودی ارزیابی می‌کنند (McMillan et al., 2002, cited by Brookhart, 2010). از آنجایی که هدف‌های آموزشی متنوع هستند این هدف‌های متنوع آموزشی در سطوح مختلف تحصیلی به وسیله طبقه‌بندی (تکسونومی) های هدف‌های آموزشی به‌نوعی سازمان و نظم دست‌یافته‌اند (Anderson et al., 2001). یکی از ابزارهای دستیابی به هدف‌های آموزشی در سنجش یادگیری دانش‌آموزان و دانشجویان، تهیه جدول مشخصات آزمون<sup>۱</sup> است. جدول مشخصات آزمون با تعیین آنچه باید در مورد اهداف یادگیری در هر حوزه محتوایی ارزیابی شود، با کمک به دریافت نمونه معرفی از محتوای موضوع درسی و اهداف یادگیری، نتایج یک ارزیابی معتبر را تضمین می‌کند (Bridge et al., 2003). هدف‌هایی که بر مبنای رفتار قابل مشاهده طراحی شده‌اند و این رفتارهای قابل مشاهده از طریق فعل‌های رفتاری (Mager, 1984, 1961) از جانب معلمان موردسنجش قرار می‌گیرند؛ اهمیت این فعل‌های رفتاری به حدی است که نویسندگان مختلف از جمله میگر در کتاب خود با عنوان "ندارک هدف‌های آموزشی" (۱۹۶۲)، ترجمه زنگنه و شمشیری، ۱۳۵۵؛ ویرایش دوم، ۱۹۸۸) فعل‌هایی که به اعمال و رفتار قابل اندازه‌گیری اشاره می‌کنند را نام می‌برد.

از آنجایی که یکی از ارکان جدول مشخصات آزمون طبقه‌بندی‌های هدف‌های رفتاری هستند، توصیفی از آن‌ها به صورت مجمل به صورت ذیل است؛ معروف‌ترین طبقه‌بندی‌های

---

1. the table of test specification

آموزشی به نام طبقه‌بندی بلوم است (Bloom et al., 1956). در این طبقه‌بندی هدف‌های آموزشی در سه دسته کلی حیطه شناختی، عاطفی و روانی حرکتی تقسیم شده‌اند (سیف، ۱۳۹۵). در نظریه یادگیری معنی‌دار Ausubel (1968) به معلمان توصیه می‌شود تا ابتدا به ارائه مقدمه‌ای از اطلاعات جدید بپردازند تا جایگاه و مکان این اطلاعات در ذهن و ساختار شناختی فراگیر مشخص شود. طبقه‌بندی Biggs and Collis (1982) تحت عنوان طبقه‌بندی سولو (ساختار نتایج یادگیری) که یک طبقه‌بندی سلسله‌مراتبی از مهارت‌های فکری است بر تعداد عناصر و روابط میان آن‌ها که دانش‌آموز باید درباره‌شان فکر کند تمرکز دارد.

Merrill (1983) اظهار می‌دارد که عملکرد برای اهداف مختلف می‌تواند دارای سطوح مختلفی باشد؛ طبقه‌بندی مریل دارای دو بعد محتوا و نوع عملکرد مورد انتظار است (Spector et al., 2005؛ به نقل از نوروزی و رضوی، ۱۳۹۰). Gagné (1985) اهداف یادگیری را در ۵ طبقه اصلی قابلیت انسانی قرار می‌دهد که شامل مهارت‌های ذهنی<sup>۱</sup>، راهبردهای شناختی<sup>۲</sup>، اطلاعات کلامی<sup>۳</sup>، مهارت‌های حرکتی<sup>۴</sup> و نگرش‌ها<sup>۵</sup> هستند.

Anderson و همکاران (2001) بازبینی را بر روی طبقه‌بندی بلوم انجام دادند؛ عمده تفاوت بین طبقه‌بندی بازبینی شده و طبقه‌بندی اصلی این است که نسخه ۲۰۰۱ دارای دو بعد دانش و فرآیندهای شناختی<sup>۶</sup> است؛ بعد دانش در طبقه‌بندی تجدیدنظرشده اندرسون و همکاران وی شامل دانش امور واقعی<sup>۷</sup>، دانش مفهومی<sup>۸</sup>، دانش روندی<sup>۹</sup> و دانش فراشناختی<sup>۱۰</sup> است. دانش واقعی شامل "عناصر اساسی است که دانش‌آموزان باید بدانند تا با یک رشته آشنا شوند یا مشکلی را در آن حل کنند" (ص ۲۹). دانش مفهومی شامل "روابط متقابل بین عناصر اساسی در ساختار بزرگ‌تر است که آن‌ها را قادر می‌سازد تا با هم عمل کنند" (ص ۲۹). دانش رویه‌ای شامل "چگونگی انجام کاری، روش‌های تحقیق و معیارهای استفاده از مهارت‌ها، الگوریتم‌ها، تکنیک‌ها و روش‌ها" است (ص ۲۹). دانش فراشناختی شامل "دانش شناخت به‌طور کلی و همچنین دانش شناخت خود" است (ص ۲۹). بعد دوم به‌عنوان حوزه

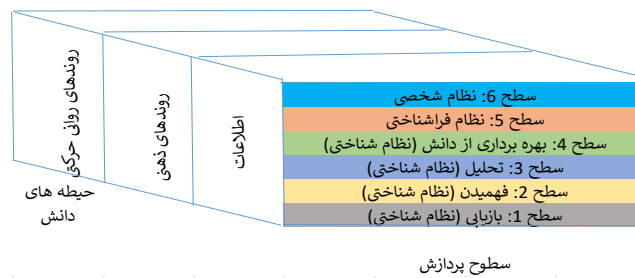
- 
1. intellectual skills
  2. cognitive strategies
  3. verbal information
  4. motor skills
  5. attitude
  6. cognitive process
  7. factual knowledge
  8. conceptual knowledge
  9. procedural knowledge
  10. metacognitive knowledge

فرآیند شناختی شناخته می‌شود و شامل شش نوع تفکر است. یادآوری<sup>۱</sup> (شامل بازیابی<sup>۲</sup> دانش مربوطه از حافظه بلندمدت<sup>۳</sup>) (ص ۳۱)؛ فهمیدن<sup>۴</sup> (شامل ایجاد "معنا از پیام‌های آموزشی، از جمله ارتباطات شفاهی، نوشتاری و گرافیکی")؛ به کار بستن<sup>۵</sup> (شامل انجام یا استفاده از "رویه‌ای در موقعیت معین")؛ تجزیه و تحلیل کردن<sup>۶</sup> (شامل تقسیم مواد به اجزای تشکیل دهنده و تعیین "نحوه ارتباط قطعات با یکدیگر و با ساختار یا هدف کلی")؛ ارزشیابی کردن<sup>۷</sup> (شامل "قضاوت بر اساس معیارها و استانداردها") و خلق کردن<sup>۸</sup> (شامل کنار هم قرار دادن "عناصر برای تشکیل یک کل منسجم یا کاربردی" و سازمان‌دهی مجدد "عناصر در یک الگو یا ساختار جدید") است (ص ۳۱). از دیدگاه طبقه‌بندی شش وجهی<sup>۹</sup> Wiggins و همکاران (2005) درک واقعی زمانی صورت می‌گیرد که دانش آموز بتواند آن را توضیح دهد<sup>۱۰</sup>، تفسیر کند<sup>۱۱</sup>، به کار بگیرد<sup>۱۲</sup>، دیدگاه داشته باشد<sup>۱۳</sup>، کند و خودشناسی<sup>۱۴</sup> داشته باشد. از نظر ایشان در آموزش، درک کامل و بالغ به‌طور ایدئال شامل توسعه کامل هر شش نوع درک است <

Anderson و همکاران (2007) Marzano and Kendall مانند Anderson و همکاران (2001) دانش را از انواع تفکر متمایز ساختند. ایشان سه حوزه دانش را اطلاعات، روندهای ذهنی و روندهای روانی حرکتی معرفی کردند طبقه‌بندی Marzano and Kendall (2007) دارای مبنای طبقه‌بندی سطح هوشیاری است و هرچه از طبقات پایین‌تر به سمت طبقات بالایی حرکت شود سطح هشیاری و سطح کنترل ذهنی موردنیاز جهت انجام یک فعالیت یا تکلیف یادگیری افزایش می‌یابد. این مدل سه نظام ذهنی را ارائه می‌دهد: نظام خود، نظام فراشناختی و نظام شناختی. چهارمین جزء مدل، دانش است. در این مدل، تکلیف جدید به‌عنوان فرصتی برای تغییر هر

- 
1. Remembering
  2. retrieving
  3. understanding
  4. Applying
  5. Analyzing
  6. Evaluating
  7. Creating
  8. six facets
  9. explain
  10. interpret
  11. apply
  12. have perspective
  13. empathize
  14. self-knowledge

کاری که فرد در زمان خاصی انجام می دهد یا به آن توجه می کند، تعریف شده. شکل زیر نظام طبقه بندی هدف های آموزشی Marzano and Kendall (2007) را نشان می دهد:



این طبقه بندی دارای دو بعد است؛ یک بعد دارای ۶ طبقه از فرآیندهای ذهنی است و بعد دیگر بعد دانش که دارای سه حیطه است. اصل زیربنایی این طبقه بندی میزان کنترل بر عملیات ذهنی است؛ به عبارتی هرچه از سطح اول به بالا حرکت می شود میزان کنترل هوشیارانه بر فرآیندهای ذهنی افزایش می یابد (به نقل از درتاج و کردنوقایی، ۱۳۹۶). ردیف های سمت راست سه نظام فکری هستند و در مورد نظام شناختی چهار جزء فرعی آن نظام نشان داده شده است. ستون های نشان داده شده در سمت چپ سه نوع یا حوزه مختلف دانش را نشان می دهد: اطلاعات، فرآیندهای ذهنی و فرآیندهای روانی حرکتی. اگر معلم مایل باشد که دانش آموز خود را در مورد یک تکنیک ویرایش خاص در عمل مشارکت دهد، گفته می شود، مثال به حوزه فرآیندهای ذهنی می پردازد. اگر معلم تربیت بدنی مایل باشد دانش آموز را درگیر فعالیت های کششی کند، مثال به حوزه فرآیندهای روانی حرکتی مربوط می شود (Marzano & Kendall, 2007). Fink (2013) طبقه بندی جدیدی از یادگیری معنی دار را توصیف می کند که طیف گسترده ای از انواع مختلف یادگیری را در برمی گیرد.

اکنون با ذکر این مطالب سؤال مهمی که به ذهن متبادر می گردد این است که چگونه آموزش های خود را به سمت سطوح مهارتی سوق دهیم؟ طبقه بندی های هدف های یادگیری که ذکر شد، ابزاری هستند که مسیر را برای دستیابی به آموزش های سطوح بالا فراهم می آورند. همچنین معلمان نمی توانند تمام موضوعات یا اهداف را اندازه گیری کنند و نمی توانند هر سؤالی را که مایل اند بپرسند. جدول مشخصات آزمون با قابلیت های خود به معلمان اجازه می دهد که یک آزمون را با تمرکز بر حوزه های کلیدی و وزن های متفاوت برای هر حوزه با توجه به اهمیت آن بسازند (Bruce, 2013)؛ ترجمه سعدی پور و همکاران، ۱۳۹۵.

یکی از نقش‌های جدول مشخصات آزمون در روایی محتوا است؛ همچنین هنگامی که در مورد وزن مؤلفه‌های سنجش تصمیم‌گیری می‌شود، این مسئله دوباره در مورد محتوا مطرح می‌شود (کیامنش و گنجی، ۱۳۸۳)؛ اما علی‌رغم اهمیت موضوع طبقه‌بندی هدف‌های یادگیری و یا استفاده از ابزاری تحت عنوان جدول مشخصات آزمون پژوهش‌های اندکی در داخل کشور و یا حتی در خارج کشور در حوزه جدول مشخصات آزمون صورت پذیرفته است. جهت بررسی پژوهش‌های انجام‌شده در حوزه مرتبط با موضوع، با یک جستجوی ساده در موتورهای جستجوی علمی داخلی، فقدان حضور یک پژوهش در خصوص نقش، کاربرد و فواید یک جدول مشخصات آزمون کاملاً محرز است. از طرفی وقتی به یک جستجوی ساده در خصوص جدول مشخصات آزمون می‌پردازیم حضور آن در حوزه پزشکی بیشتر جلوه می‌نماید و در سایر حوزه‌های علم خیلی به آن پرداخته نشده است؛ اما در میان پژوهش‌های خارجی نیز بیشتر پژوهش‌هایی که صورت گرفته است به نقش جدول مشخصات آزمون در حوزه علوم پزشکی پرداخته است. از جمله Patel همکاران (2016) نشان داده‌اند که استفاده از یک بلوپرینت آزمون سبب وزن‌دهی مناسب سؤالات برای موضوعات مختلف می‌شود (۹۰٪)، سؤالات متناسب با اهداف یادگیری طراحی می‌شود (۸۰٪) و از طراحی افراطی سؤالات بسیار آسان یا بسیار دشوار جلوگیری می‌شود. یا Eweda و همکاران (2020) نشان دادند که استفاده از بلوپرینت آزمون به اطمینان از کیفیت و اینکه میزان روایی محتوا در ارزیابی دانش‌آموزان افزایش می‌یابد می‌پردازند. حتی در برخی پژوهش‌ها ادعا شده است که بین آموزش و یادگیری تعادل ایجاد می‌کند، روایی، پایایی و مقبولیت ارزیابی‌ها را نیز ارتقا می‌دهد (Kaur et al., 2021). مطالعه گیل و سن نشان داد که در بررسی مطالعات بر اساس رویکرد طبقه‌بندی تجدیدنظر شده بوم ۸ درصد از مطالعات در سطح یک بوم، ۲۰ درصد سطح ۲ و ۸ درصد در سطح ۳ بودند؛ در حالی که ۶۴ درصد از سطح سؤالات مبهم بودند (Gill, & Sen, 2020). Patil و همکاران (2015) معتقدند که آزمون‌های کتبی معایب متعددی از جمله تعداد سؤالات، نمونه‌گیری محدود؛ ناعادلانه توزیع سؤالات در موضوعات و سؤالات مبهم دارند؛ بدین منظور در پژوهش خود از یک بلوپرینت استفاده کردند که نتایج به این صورت بود که همه اعضای هیئت علمی معتقد بودند که بلوپرینت‌ها ارزیابی را با اهداف آموزشی هماهنگ می‌کرد و به‌عنوان راهنما کمک‌کننده

بود؛ بنابراین خلأ بررسی نقش جدول مشخصات آزمون در پیشینه پژوهش کاملاً مشخص است.

علاوه بر موارد مذکور با توجه به توسعه پرسرعت تکنولوژی، نقش آن در حوزه سنجش عملکرد دانش‌آموزان اجتناب‌ناپذیر است؛ طراحی سیستم‌های جدید ارزیابی با استفاده از فناوری‌های جدید می‌بایست بر یافتن راه‌هایی برای ارزیابی دقیق و در حوزه اهداف آموزشی متمرکز شود. هدف پژوهش حاضر آن است که نرم‌افزار این قابلیت را داشته باشد که بنا بر تعاریف و الگوریتم‌هایی که بر مبنای افعال طراحی شده‌اند سنجش را به سمت سطوح بالای شناختی سوق دهد. در این راستا دو رویکرد جدیدتر (طبقه‌بندی تجدیدنظر شده اندرسون و همکاران و طبقه‌بندی مارزانو و کندال) مدنظر این پژوهش قرار دارد. دو طبقه‌بندی اندرسون و همکاران وی و مارزانو و کندال دارای رویکرد سلسله‌مراتبی هستند؛ بنابراین در این پژوهش به تلفیق دو رویکرد سلسله‌مراتبی پرداخته و جدول مشخصاتی بر مبنای آن طراحی شد. همچنین با توجه به اهمیت درس علوم و نقش آن در شکل‌گیری درس‌های زیست‌شناسی، زمین‌شناسی، فیزیک و شیمی در مقاطع بالاتر دبیرستان و رشته‌های علوم پایه در دانشگاه، به‌عنوان درس موردنظر جهت طراحی جدول مشخصات آزمون انتخاب شد؛ به‌طوری‌که معلمان در طراحی سؤالات آزمون پیشرفت تحصیلی درس علوم بتوانند از این ابزار استفاده کنند.

سؤال‌های پژوهش از قرار زیر است:

۱. ویژگی‌های یک ابزار حمایتی معتبر و عملی که پتانسیل تأثیرگذاری بر عملکرد معلمان در توانمندسازی آن‌ها برای طراحی سؤالات با کیفیت در سنجش سطوح بالای شناختی را داراست، چیست؟
۲. آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های طراحی سؤالات بر مبنای سطوح بالای شناختی را برطرف کند؟
۳. آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های ارتقای توانمندی معلمان در طراحی سؤالات با کیفیت را برطرف کند؟

این سؤال شامل موارد زیر است:

۱. مخاطبان هدف چه کسانی باید باشند؟
۲. این ابزار باید از چه طریقی در دسترس مخاطبان باشد؟

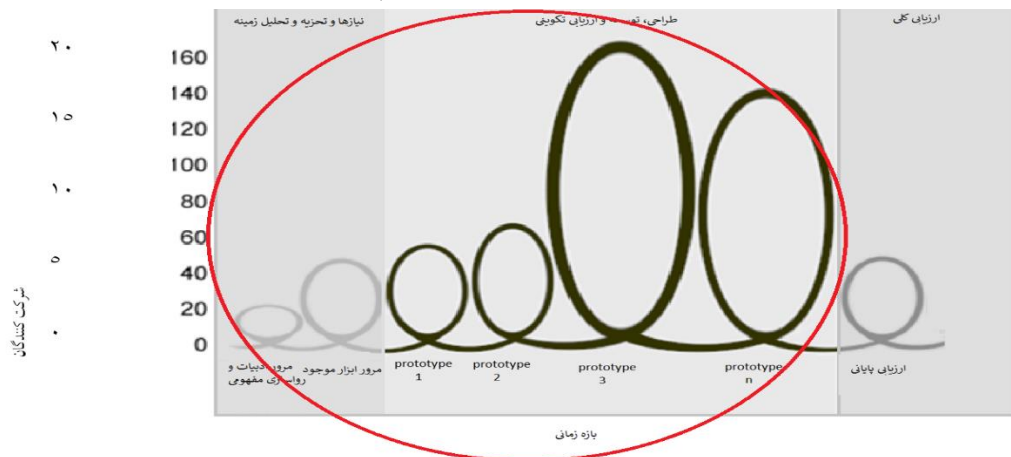


۳. این ابزار از چه جنبه‌هایی باید از توسعه ارزیابی حمایت کند؟

### روش

پژوهش حاضر از نوع توسعه‌ای بود؛ شکل ۱، سه مرحله اصلی مطالعه (تجزیه و تحلیل، طراحی و ارزیابی) را نشان می‌دهد. از آنجایی که مقاله حاضر برگرفته از رساله دکتری با همین موضوع است، بخش طراحی و توسعه نرم‌افزار بر اساس رویکرد ترکیبی و با چشم‌انداز غالب معماری نرم‌افزار بر آن حاکم است؛ لذا بر مبنای شکل موردنظر، شامل مراحل یک و دو یعنی مراحل نیازها و تجزیه و تحلیل زمینه و طراحی، معماری و ارزیابی تکوینی می‌شود. همچنین با توجه به اینکه هدف تهیه ابزاری مبتنی بر وب جهت طراحی آزمون است؛ لذا تعریفی از جامعه نمی‌توان برای آن ارائه داد. از طرفی چون ابزاری مبتنی بر شبکه است جامعه را می‌توان کلیه ابزارهای آزمون‌سازی تحت وب دانست. جامعه مورد مطالعه متخصصان جهت توسعه، بررسی روایی و اعتبار نرم‌افزار شامل دو نوع شرکت‌کننده گروه‌های کاربری و گروه‌های خبره بود. گروه کاربری شامل معلمان و گروه‌های تخصصی متشکل از اساتید دانشگاه فرهنگیان در حوزه آموزش ابتدایی به‌ویژه آموزش علوم، متخصصان سنجش و اندازه‌گیری و کارشناسان نرم‌افزار که به‌صورت هدفمند انتخاب شدند.

شکل ۱. مراحل اصلی و فرعی در توسعه نرم‌افزار



## یافته‌ها

این بخش با رویکرد پاسخ به سؤالات پژوهش بر اساس منطق حاکم بر رویکرد تحقیق و توسعه طراحی شد:

سؤال اول پژوهش: ویژگی‌های یک ابزار حمایتی معتبر و عملی که پتانسیل تأثیرگذاری بر عملکرد معلمان در توانمندسازی آن‌ها برای طراحی سؤالات با کیفیت در سنجش سطوح بالای شناختی را داراست، چیست؟

مرحله اول تحقیق و توسعه:

۱- مرحله تحلیل نیازها و تجزیه و تحلیل زمینه: برای پیشبرد مرحله اول (تجزیه و تحلیل) در ابتدا به مرور تفصیلی ادبیات پژوهش پرداخته شد؛ همچنین نرم‌افزارهایی که در زمینه طراحی سؤال، پرسشنامه و آزمون به صورت تخصصی فعالیت دارند بررسی شدند؛ یک چک‌لیست مقایسه‌ای از قابلیت‌های نرم‌افزار به روز دنیا که قابلیت‌های هر کدام را نشان می‌دهد و در جدول ۱ نشان داده شده است.

بنابراین در پاسخ به سؤال اول مبنی بر آنکه ویژگی‌های ابزاری که امکان سنجش سطوح بالای شناختی را فراهم می‌سازد می‌توان به ابزاری تحت عنوان نرم‌افزار کاربردی جدول مشخصات آزمون اشاره کرد که ضمن دارا بودن قابلیت‌هایی جهت نمایه طبقات شناختی بر اساس رویکردهای مختلف طبقه‌بندی‌های هدف‌های رفتاری به صورت بصری برای کاربران و محتوای موردنظر، قابلیت وزن‌دهی به هر یک طبقات و محتوای موردنظر را فراهم می‌سازد؛ از این جهت امکان سوق دادن سنجش به سمت سطوح بالای شناختی را از جهت تعیین طبقات، فعل‌هایی که به سنجش هر یک از سطوح می‌پردازند و هدف‌های رفتاری دقیق، فراهم می‌سازد؛ همچنین چنین ابزاری با کمک به دریافت نمونه معرفی از محتوای موضوع درسی و اهداف یادگیری، نتایج یک ارزیابی معتبر را تضمین می‌کند.

سؤال دوم پژوهش: آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های طراحی سؤالات بر مبنای سطوح بالای شناختی را برطرف کند؟ جهت پاسخ به این سؤال ابتدا بر مبنای یافته‌های مطالعه مبانی نظری و ابزارهای موجود، الگوی اولیه نرم‌افزار طراحی شده معرفی می‌شود؛ سپس به قابلیت‌هایی که یک نرم‌افزار می‌تواند داشته باشد که امکان دستیابی به ارزیابی سطوح بالای شناختی را فراهم کند اشاره می‌شود.

جدول ۱. چک‌لیست مقایسه قابلیت‌های نرم‌افزارها

	Kahoot!	surveyplanet	qualtrics	Google Forms	sogolytics	surveylegend	surveysparrow	questionpro
امکان طراحی جدول مشخصات آزمون با تنوع نوع سوال	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
امکان محاسبه شاخص‌های روانسنجی (ضریب تمیز، ضریب دشواری، و ...)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
امکان استفاده از داده‌های آزمون برای ارزشیابی آغازین، تکوینی و پایانی	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
امکان تعیین وزن و تعداد سوالات بر اساس زمان و اهمیت موضوعات درسی	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
امکان تولید گزارش‌های تحلیلی و گرافیکی از نتایج آزمون	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
امکان به اشتراک‌گذاری و همکاری با دیگران در طراحی و اجرای آزمون	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
امکان سفارشی‌سازی ظاهر و رابط کاربری آزمون	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

۲- مرحله طراحی، معماری و ارزیابی تکوینی: در ادامه روند توسعه نرم‌افزار، از طراحی الگوها و چرخه‌های تشکیل‌دهنده هر یک به صورت مسیری چرخشی و دورانی ادامه می‌یابد که ترسیم پیش‌بینی مراحل این روند مسیر زیر را به دست می‌دهد:

الگوی<sup>۱</sup> اولیه (پروتوتایپ) (۱): اولین الگوی اولیه نرم‌افزار بر اساس رویکرد تلفیقی طبقه‌بندی‌های هدف‌های یادگیری طراحی شد. شمای اولیه نرم‌افزار طراحی شد و با توجه به حیطه‌ها و زیرمجموعه‌های هر یک از حیطه‌ها، طبقه‌بندی‌ها با جزئیات طراحی شدند (شکل ۱).

شکل ۲. طبقه‌بندی‌های هدف‌های رفتاری

طبقه بندی ادغام شده		طبقه بندی اندرسون و همکاران		طبقه بندی مارزانو و کندال		
نظام شناختی	بازیابی	یادآوری	بازیابی	نظام شناختی	فهمیدن (درک)	
	فهمیدن	فهمیدن	فهمیدن (درک)		تحلیل	
	به کار بستن	به کار بستن	تحلیل کردن		بهره برداری از دانش	
	تحلیل	تحلیل کردن	ارزشیابی کردن	نظام فراشناختی	مشخص کردن اهداف	
	بهره برداری از دانش	ارزشیابی کردن	آفریدن		نظارت بر فرآیند	
نظام خود	ارزشیابی کردن	آفریدن	آفریدن	نظام خود	نظارت بر روشنی	
	آفریدن	مشخص کردن اهداف	نظارت بر صحت		آزمودن اهمیت	
	مشخص کردن اهداف	نظارت بر فرآیند	نظارت بر روشنی	آزمودن کارآمدی	نظام خود	آزمودن پاسخ‌های هیجانی
	نظارت بر فرآیند	نظارت بر روشنی	نظارت بر صحت	آزمودن سطح انگیزش		

در مرحله بعد، برای طبقه‌بندی ترکیبی اندرسون و همکاران و مارزانو و کندال با طی مراحل زیر، هدف‌های رفتاری تعریف گردید؛ در ابتدا فعل‌های رفتاری که از منابع نظری برای هر یک از طبقات گردآوری شده بودند مورد بررسی قرار گرفتند؛ بدین ترتیب تعداد فعل‌های اولیه ۱۲۷ فعل به دست آمد. در این مرحله با استفاده از امکان هوش مصنوعی chatGPT-4 مراحل زیر طی شد:

۱. هر فعل به صورت جداگانه وارد chatGPT-4 شد و از آن خواسته شد که معادل‌های انگلیسی فعل مورد نظر را تولید کند؛ پس از بررسی در دیکشنری و راستی‌سنجی معنی فعل، آن فعل وارد لیست فعل‌های مورد نظر برای تولید هدف‌های رفتاری در رویکرد ترکیبی شد. به عنوان نمونه فعل‌های زیر، فعل‌هایی بودند که به عنوان فعل‌های معادل توسط chatGPT-4 به عنوان فعل‌های معادل تولید شدند؛ بدین ترتیب تعداد فعل‌ها به ۴۰۶ فعل افزایش پیدا کرد.

بازشناسی (شناسایی می‌کند) to detect

بازشناسی می‌کند (شناسایی می‌کند) to acknowledge

بازشناسی می‌کند (شناسایی می‌کند) to discern

بازشناسی می‌کند (شناسایی می‌کند) to perceive

بازشناسی می‌کند (شناسایی می‌کند) to spot

بازشناسی می‌کند (شناسایی می‌کند)، to identify

بازشناسی می‌کند (شناسایی می‌کند)، to recognize

بازشناسی می‌کند (شناسایی می‌کند)، to realize

۲. در مرحله بعد از chatGPT-4 خواسته شد تا تعاریف، شباهت، تفاوت و مثال برای هر یک از فعل‌ها تولید کند (به این ترتیب یک معنای general برای فعل‌ها تولید شد). پس از آن از chatGPT-4 خواسته شد تا کاربرد فعل مورد نظر در حیطه آموزش و یادگیری را توضیح دهد؛ همه این تولیدات در فایل پیش فرض نرم‌افزار (اولیه) ذخیره شده است تا در هنگام استفاده معلم پیش از استفاده از فعل و تولید سؤال برای او به نمایش در آید.

۳. در مرحله بعد با ارائه تعاریف داده شده به chatGPT-4 از آن خواسته شد تا مراحل انجام فعل مورد نظر را با جزئیات و به صورت مرحله‌ای تولید کند.

۴. در نهایت با تهیه یک پرامپت (prompt) از chatGPT-4 خواسته شد تا برای هریک از فعل‌ها برای هریک از طبقات مورد نظر هدف رفتاری یک مثال تولید کند تا یک راهنمای دقیق برای تولید سؤال در اختیار معلم قرار داشته باشد.

پس از آنکه مراحل انجام فعل با توجه به توضیحات فوق مشخص شد، با ارائه مراحل به chatGPT-4 و تولید یک پرامپت از آن خواسته شد که یک مثال به عنوان تمرین تولید کند. مثال زیر برای فعل (مثال غیر معمول می‌زند) به شرح ذیل است:

\* هدف رفتاری: دانش‌آموز قادر خواهد بود مطابق با دستورالعمل زیر، در مورد واژگان درس علوم مثال غیر معمول بزند.

مراحل مثال غیر معمول زدن to give an atypical example یک موضوع می‌تواند به صورت زیر باشد:

۱. تعریف موضوع: ابتدا، باید موضوعی که قرار است مثال غیر معمولی برای آن ارائه شود، تعریف کنید. این می‌تواند هر چیزی از یک گروه خاص تا یک مفهوم عمومی باشد.

۲. فهم فرضیات رایج: سپس باید درک کافی از فرضیات یا ایده‌های رایج درباره این موضوع داشته باشید. این می‌تواند به وسیله تحقیق، مطالعه، یا مشاهده به دست آید.

۳. انتخاب مثال غیر معمول: در این مرحله، شما باید یک مثال انتخاب کنید که معمولاً در توصیف این موضوع یا گروه بکار نمی‌رود. این مثال باید مناسب و معتبر باشد، ولی در عین حال از ایده‌های رایج چالش کند.

۴. توضیح مثال: بعد، شما باید مثال خود را به صورت دقیق و کامل توضیح دهید، نشان دهید که چگونه از فرضیات رایج فراتر می‌رود.

۵. تأکید بر چرایی غیر معمول بودن: در نهایت، باید تأکید کنید که چرا مثال شما غیر معمول است و چگونه می‌تواند درک ما از موضوع را تغییر دهد. این بخش می‌تواند شامل یک بررسی انتقادی از فرضیات رایج باشد.

مثال:

دانش‌آموز قادر خواهد بود با استفاده از دستورالعمل ذکر شده، در مورد واژگان درس

علوم در تمرین "مثال غیر معمول در درک فتوسنتز" مثال غیر معمول بزند.

تعریف موضوع: موضوع این تمرین، فتوسنتز است. فتوسنتز فرایندی است که در آن گیاهان و برخی از سایر ارگانیسم‌ها، از طریق استفاده از نور خورشید، انرژی نور را به انرژی شیمیایی تبدیل می‌کنند و از دی‌اکسید کربن و آب، گلوکز و اکسیژن تولید می‌کنند. فهم فرضیات رایج: قبل از ارائه مثال غیرمعمول، دانش‌آموز باید درک کافی از فرضیات رایج درباره فتوسنتز داشته باشد. این می‌تواند شامل مفهوم و مراحل فتوسنتز، نیاز به نور، و استفاده از دی‌اکسید کربن و آب باشد.

انتخاب مثال غیرمعمول: مثال غیرمعمول برای فتوسنتز این است که برخی گیاهان، در شرایط خاصی قادر به انجام فتوسنتز بدون نیاز به نور هستند؛ به عبارت دیگر، این گیاهان توانایی تولید گلوکز و اکسیژن را در شب و در تاریکی کامل دارند.

توضیح مثال: این گونه گیاهان، که به آن‌ها گیاهان C4 گفته می‌شود، در برخی مناطق گرم و خشک که دما بسیار بالا است، رشد می‌کنند. برای جلوگیری از تبخیر زیاد آب در روزهای گرم، این گیاهان در طول روز، سوراخ‌هایی را در برگ‌های خود می‌بندند تا آبی که درون آن نگه داشته‌اند، تبخیر نشود؛ اما در شب، این سوراخ‌ها را باز کرده و فتوسنتز را انجام می‌دهند.

تأکید بر چرایی غیرمعمول بودن: این مثال غیرمعمول نشان می‌دهد که علم و فرضیات رایج همواره قابل چالش هستند. چراکه در اکثر گیاهان، نیاز به نور برای فتوسنتز ضروری است، ولی گیاهان C4 در شرایط خاصی توانایی انجام فتوسنتز در تاریکی را دارند. این نمونه نشان می‌دهد که همه جوانب یک موضوع را باید مورد بررسی قرار داد و به دیدگاه‌های جدید و غیرمعمول هم توجه کرد تا در کمان از موضوعات پیشرفت کند.

با استفاده از این روش در نهایت تعداد ۵۱۳۷ هدف رفتاری برای طبقه‌بندی ترکیبی تولید شد. در نهایت چک‌لیستی تهیه شد که از خبرگان پرسیده شد که آیا فعل در طبقه مورد نظر قرار می‌گیرد یا خیر؛ و در صورتی پاسخ منفی در کدام طبقه قرار می‌گیرد؛ که البته با توجه به آنکه هسته اولیه فعل‌ها از مبانی نظری انتخاب شده بود تغییری در طبقه‌بندی فعل‌ها رخ ندارد (تأیید روایی طبقه) و نظر خبرگان بر هدف‌های رفتاری تولیدشده نیز مساعد بود.

در مرحله دوم با مطالعه مبانی برنامه‌نویسی و کسب اطلاعات از کارشناسان نرم‌افزار الگوریتم طراحی نرم‌افزار به صورت مرحله‌ای طراحی شد. با توجه به وسعت کار تولیدشده

امکان نمایش تمام نمای آن در این فضا موجود نیست؛ اما جهت نمایش بخشی از کار، تصویر زیر نشان داده می‌شود:

ردیف	اطلاعات (دانش امور واقعی، دانش مفهومی)		روندهای ذهنی (دانش روندی):				روندهای روانی حرکتی:	
	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
3	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
4	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
5	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
6	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
7	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
8	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
9	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
10	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
11	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
12	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
13	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
14	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
15	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
16	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
17	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
18	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
19	دانش	دانش	دانش	دانش	دانش	دانش	دانش	
20	دانش	دانش	دانش	دانش	دانش	دانش	دانش	

در این مرحله پس از طراحی شمای ابتدایی نرم‌افزار، از معلمان در خصوص پیشنهادهای ایشان سؤال شد و پس از دریافت نظرات، ایده‌هایی دریافت شد که در طراحی نمونه اولیه امکان اعمال برخی فراهم گردید.

سؤال سوم پژوهش: آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های ارتقای توانمندی معلمان در طراحی سؤالات با کیفیت را برطرف کند؟ این سؤال شامل: (مخاطبان هدف چه کسانی باید باشند؟ این ابزار از چه طریقی در دسترس مخاطبان باشد؟ و این ابزار از چه جنبه‌هایی باید از توسعه ارزیابی حمایت کند؟) با توجه به توضیحاتی که پیش‌تر داده شد مخاطبان هدف این نرم‌افزار تمام معلمان (درس علوم) هستند. جهت پاسخ به این سؤال که این نرم‌افزار از چه طریقی در اختیار مخاطبان قرار می‌گیرد؟ در بخش معماری نرم‌افزار به این موضوع پرداخته شده است که نرم‌افزار در قالب وب‌اپ در اختیار کاربران قرار گیرد؛ تا کاربران با انواع ابزارها اعم از گوشی، لپ‌تاپ، تبلت و کامپیوترها با انواع سیستم‌عامل‌ها به آن دسترسی داشته باشند؛ و در پاسخ به این سؤال که این نرم‌افزار از چه جنبه‌هایی از توسعه حمایت می‌کند؟ به این موضوع اشاره شده است که روش‌های زیر می‌تواند به توسعه ارزیابی توسط معلم حمایت کند:



۱- هنگام استفاده از نرم‌افزار امکان تخصیص نسبت‌های موردنظر برای فصل‌ها یا محتوای موردنظر را فراهم می‌کند؛ بنابراین معلم دقیقاً تعیین می‌کند که بر اساس زمان، محتوا یا اهمیت، نسبتی را به محتوای موردنظر جهت طراحی سؤال در نظر بگیرد.

۲- طبقات شناختی و غیرشناختی به معلم نمایش داده می‌شود و او می‌تواند حتی برای سطوح غیرشناختی (نظام خود و فراشناختی) نیز سؤال طراحی کند.

۳- سطوح مختلف طبقات شناختی به معلم نشان داده می‌شود و او ملزم می‌شود که تنها به اندازه‌گیری سطوح بازیابی یا فهمیدن نپرداخته و از بین سطوح بالاتر نیز انتخاب کند و نسبتی را برای هر یک از طبقات در طراحی سؤالات در نظر بگیرد.

۴- فعل‌های رفتاری مربوط به هر طبقه پس از انتخاب طبقه موردنظر برای او به نمایش درمی‌آید و او قادر است از طیف متنوعی از فعل‌ها جهت طراحی سؤالات استفاده کند؛ حتی این امکان برای این ابزار طراحی شده است که برای فعل‌هایی که در زبان انگلیسی چندین معادل برای آن‌ها وجود دارد نیز امکان طراحی سؤال وجود داشته باشد (به‌عنوان مثال فعل بازشناسی کردن یا انتخاب کردن که دارای انواع فعل‌های مختلف با کارکردهای مختلف در زبان انگلیسی هستند).

۵- هدف‌های رفتاری متناظر به فعل موردنظر به همراه یک نمونه مثال مربوط به درس علوم برای او به نمایش درمی‌آید و او دقیقاً متوجه می‌شود که برای فعل بازشناسی می‌کند (to discern) از دانش آموز باید توقع چه فعالیت‌هایی را داشته باشد.

۶- امکان طراحی سؤال در قالب‌های مختلف برای معلم فراهم شده است. در نتیجه امکانات فوق‌تلاشی در جهت توسعه ارزیابی بوده است.

معماری نرم‌افزار: نرم‌افزار جدول مشخصات آزمون موردنظر این پژوهش، در قالب وب اپ طراحی شد؛ در پاسخ به چرایی انتخاب وب اپ، می‌توان این‌طور اشاره کرد که به‌صورت آنلاین و با استفاده از مرورگرهای وب در دسترس خواهد بود؛ تمامی کاربران در تمامی دستگاه‌ها (اعم از موبایل، لپ‌تاپ و پی‌سی) و تمامی سیستم‌عامل‌ها (اعم از ویندوز، لینوکس، مک و سیستم‌عامل‌های موبایل) قابلیت استفاده از آن را دارند؛ بنابراین وب اپ‌ها از آن‌جهت که از طریق مرورگر وب قابل دسترسی هستند این امکان را به کاربران خود می‌دهند که از هر دستگاهی با اتصال به اینترنت به آن‌ها دسترسی پیدا کنند؛ همچنین نیازی به نصب ندارند و با توجه به اینکه از پروتکل HTTP استفاده می‌کنند کاربران به‌صورت

آنلاین آن‌ها را در مرورگر خود اجرا می‌کنند. از دیگر ویژگی‌های این اپلیکیشن‌ها، امکان ایجاد تعامل پویا همچون انواع فرم‌ها یا قابلیت‌های تعاملی دیگر با کاربران است و همین ویژگی سبب توسعه کسب و کارها، و افزایش دسترسی به خدمات می‌شود؛ وب‌اپ‌ها امکان ذخیره اطلاعات در سرورهای آنلاین برای استفاده کاربران را فراهم می‌کنند. نکته بسیار جالب در مورد وب‌اپ‌ها آن است که به‌روزرسانی‌هایی که در نرم‌افزار انجام می‌شود به‌صورت مرکزی و در سرورها انجام شده و نتیجه آن می‌شود که کاربران همیشه امکان دسترسی به آخرین نسخه را دارند. وب‌اپ‌ها با افزونه‌های مختلف وردپرس سازگار شده و امکان ارائه تجربه بهتری را برای کاربران فراهم می‌کنند؛ همچنین امکان رعایت استانداردهای سئو (SEO) یا بهینه‌سازی برای موتورهای جستجو و سرعت بارگذاری برای بهبود عملکرد وب‌اپ فراهم می‌شود. این ویژگی‌ها سبب محبوبیت بالای وب‌ها در دنیای دیجیتال شده است.

تمامی وب‌اپ‌ها شامل دو بخش اصلی کلاینت یا فرانت‌اند (Front-end) و بخش سرور یا بک‌اند (Back-end) می‌شوند. این تقسیم‌بندی از آن جهت که وظایف مختلفی را به هر قسمت تخصیص می‌دهد در توسعه وب‌اپ‌ها بسیار پراهمیت است. کلاینت یا فرانت‌اند آن بخش از وب‌اپ است که توسط کاربران قابل مشاهده بوده و از عناصری همچون دکمه‌ها، عکس‌ها، و تمامی اجزای قابل مشاهده و تعاملی وب‌اپ تشکیل یافته است و مسئولیت نمایش و یا تعامل با کاربران را بر عهده دارد. برای پیاده‌سازی فرانت‌اند از زبان‌های برنامه‌نویسی همچون JavaScript<sup>۱</sup>، HTML<sup>۲</sup> و CSS<sup>۳</sup> استفاده می‌شود. JavaScript زبان برنامه‌نویسی است که در سمت کلاینت مورد استفاده قرار می‌گیرد و این اجازه را می‌دهد که رفتار صفحات وب را تغییر دهیم. سرور یا بک‌اند پشت‌صحنه وب‌اپ بوده و شامل سرور، پایگاه داده و برنامه‌های سرور است و مسئولیت منطق و پردازش فعالیت‌ها را بر عهده دارد. در بک‌اند می‌توان از زبان‌های برنامه‌نویسی همچون Python، PHP، Ruby، و غیره استفاده کرد. PHP با پایگاه‌های داده مختلفی همچون MySQL، SQL Server، PostgreSQL و Oracle ادغام می‌شود. همچنین با فریم‌ورک‌های گوناگونی از جمله لاراول<sup>۴</sup> کار می‌کند و از این جهت توسعه وب را ساده‌تر می‌سازد.

- 
1. Search Engine Optimization
  2. Hyper Text Markup Language
  3. Cascading Style Sheets
  4. Laravel

در برنامه‌های وب، معمولاً سایت‌ها به‌عنوان "رابط برنامه‌نویسی اپلیکیشن" API<sup>۱</sup> با استفاده از پروتکل REST با سرور ارتباط برقرار می‌کنند. API مجموعه‌ای از پروتکل‌ها، روندها و ابزارها است که به نرم‌افزارها اجازه می‌دهد با یکدیگر ارتباط برقرار کنند؛ به زبان ساده‌تر API واسط بین کلاینت و سرور است که دیتا را ارسال و دریافت می‌کند. API چگونگی تعامل بین دو نرم‌افزار را تعریف می‌کند و تحت عنوان رابط فرانت‌اند و بک‌اند عمل کرده و اجازه تعامل فرانت‌اند با بک‌اند را می‌دهد. API ها در بسیاری از جنبه‌های مختلف فناوری و نرم‌افزار از جمله توسعه وب، ارتباط با پایگاه داده، اتصال به سرویس‌های آنلاین، و ... استفاده می‌شوند. به‌منظور ارتباط بین کلاینت و سرور معمولاً از پروتکل REST<sup>۲</sup> استفاده می‌شود. پروتکل REST استانداردی برای دریافت و ارسال داده‌ها در دو بخش کلاینت و سرور است. در پژوهش حاضر نیز از پروتکل رست استفاده شد و سایت نیز به‌تبع سایت API REST خواهد بود. همچنین در پژوهش حاضر، کلاینت با استفاده از فریم‌ورک react که بر پایه زبان javascript است و دارای ویژگی‌ها منحصربه‌فرد خود است و بک‌اند با زبان PHP و فریم‌ورک<sup>۳</sup> لاراول پیاده‌سازی شد. فریم‌ورک‌ها قالب‌ها و ساختارهای ثابتی دارند که برنامه‌نویسان با استفاده از آن‌ها قادر خواهند بود برای توسعه وب اپ‌های خود از آن‌ها استفاده کنند. زبان سمت سرور زبان PHP و فریم‌ورک لاراول، و زبان سمت فرانت‌اند، زبان javascript و فریم‌ورک react خواهد بود. همچنین برای این پژوهش از پایگاه داده MySQL استفاده شد؛ MySQL یک سیستم مدیریت برای پایگاه داده است و به برنامه‌نویسان اجازه می‌دهد که داده‌های خود را در یک پایگاه داده از نوع رابطه‌ای RDBMS ذخیره و پس از آن مدیریت کنند. (مصاحبه با متخصص برنامه‌نویس، MySQL به‌عنوان یک سیستم پایگاه داده قدرتمند همراه با استفاده راحت از آن و دارا بودن توانایی‌ها و نیازمندی‌های یک وب‌سایت برای جستجو و ارائه داده‌ها به مرورگر در صدر انتخاب‌ها است. پیوند PHP و MySQL برای ذخیره و بازیابی داده‌ها، ابزارهای مهم برای توسعه وب‌سایت‌ها و قابلیت‌های لازم برای ایجاد WEB 2.0 را برای مخاطبان فراهم می‌کند (Nixon, 2012)، ترجمه جمشیدی شیخ‌آبادی، ۱۳۹۹).

الگوی اولیه (پروتوتایپ) (۲) در ادامه بررسی شده است:

- 
1. Application Programming Interface
  2. Representational State Transfer
  3. Framework

اولین چرخه الگوی دوم. ارزیابی انجام‌شده بر روی الگوی اول نرم‌افزار طراحی‌شده، مجموعه‌ای از مشخصات محصول را به همراه خواهد داشت که برای ادامه توسعه استفاده می‌شود. الگوی اولیه به کاربران کمک می‌کند تا به طراحی سؤال پردازند و سپس پرونداد را در یک واژه‌پرداز باز کرده و آنها را چاپ کنند. جهت ارزیابی نمونه اولیه برخی پیشنهادها و اصلاحات در مراحل دستیابی به جدول مشخصات نهایی داده شده، و جلساتی با متخصص نرم‌افزار جهت رفع اشکالات و کج‌فهمی‌ها در نمونه اولیه تولیدشده برگزار شد. دومین چرخه الگوی دوم. در طول غربال الگوی دو، توسعه‌دهنده یک الگوی اصلی از نرم‌افزار آماده کرد. جلساتی به صورت آنلاین و حضوری بین توسعه‌دهنده و سایر متخصصان در حوزه‌های مختلف ذکرشده برگزار شد تا ساختار طراحی‌شده را بررسی کنند. این در ابتدای چرخه الگوی دوم اتفاق می‌افتد و ایده‌های به‌دست‌آمده پس از ثبت و بازبینی، در طراحی نرم‌افزار موردتوجه قرار گرفت.

نوع تخصیص سوالات

بودجه بندی جدول مشخصات بر اساس ملاک حجم مطالب  
 بودجه بندی جدول مشخصات بر اساس ملاک اهمیت مطالب  
 بودجه بندی جدول مشخصات آزمون بر اساس ملاک زمان تدریس  
 بودجه بندی جدول مشخصات آزمون بر اساس ملاک تعداد اهداف رفتاری

نوع طبقه بندی

طبقه بندی ترکیبی (اندروسون و همکاران وی و رویکرد مارزتو و کندال)  
 طبقه بندی فینک

سطوح امتحان

جدول مشخصات ارزشیابی پیش از آزمون (آزمون آغازین- آزمون آمادگی سنجش رفتارهای ورودی) ملاکی  
 جدول مشخصات ارزشیابی پیش از آزمون (آزمون آغازین- آزمون جابایی یا سنجش میزان اطلاعات یادگیرندگان از هدف های جدید درس) هنجاری  
 جدول مشخصات ارزشیابی ضمن آزمون (آزمون تشخیصی) ملاکی  
 جدول مشخصات ارزشیابی ضمن آزمون (آزمون تشخیصی) ملاکی  
 جدول مشخصات ارزشیابی پس از آزمون (آزمون تکمیلی) ملاکی  
 جدول مشخصات ارزشیابی پس از آزمون (آزمون تکمیلی) هنجاری

یکی از پایه ها را انتخاب کنید

پایه ششم

نوع تخصیص سوالات

بودجه بندی جدول مشخصات بر اساس ملاک

بودجه بندی جدول مشخصات بر اساس ملاک حجم مطالب: بخش‌هایی از هدف و محتوا که دارای حجم بیشتری می‌باشد و زمانی بیشتری از آموزش صرف آن‌ها شده است (سیف، 1393). این ملاک زمانی کاربرد دارد که محتوا در یک یا چندین منبع موجود است و حجم مطالب را می‌توان به طرق مختلف (مثلاً تعداد صفحه یا فصل) اندازه‌گیری کرد.

نوع طبقه بندی

طبقه بندی ترکیبی (اندرسون و همکاران وی و

سطوح امتحان

بعضی‌ها مثال اگر تعداد صفحات کتاب علوم پایه ششم ابتدایی دارای 103 صفحه است و تعداد صفحات درس 13 (سالم بنامیم) 7 صفحه است این بخش از محتوا از 20 سؤال دارای تعداد سؤالات زیر خواهد بود:  
 $1.3 = 20 \times 7 / 103$  سؤال

با این روش حجم کل مطالب و حجم هر بخش از محتوا به صورت جداگانه‌ای مشخص شده و برای تعیین سهم هر بخش محتوا، حجم مطالب بخشی که مورد نظر است بر کل محتوا تقسیم می‌شود (ایرملی، ایزدخواستی و اسماعیلی، 1397).

## بحث و نتیجه‌گیری

در این بخش سؤالات پژوهش مجدداً طرح شده و به بحث پرداخته می‌شود؛ سؤال اول پژوهش: ویژگی‌های یک ابزار حمایتی معتبر و عملی که پتانسیل تأثیرگذاری بر عملکرد معلمان در توانمندسازی آن‌ها برای طراحی سؤالات با کیفیت در سنجش سطوح بالای شناختی را داراست، چیست؟

جهت پاسخ به سؤال اول، پژوهش حاضر به بررسی ادبیات و نرم‌افزارهای موجود در زمینه طراحی سؤالات، به دنبال شناسایی ویژگی‌های یک ابزار کارآمد برای کمک به معلمان در سنجش سطوح بالای شناختی بود. این ابزار با بهره‌گیری از جدول مشخصات آزمون، قابلیت انتخاب یک نمونه معرف از محتوا را دارد. همچنین با توجه به تفاوت‌های فردی دانش‌آموزان و تقویت مهارت‌های اندیشیدن، کاوشگری و خلاقیت آن‌ها، سعی خواهد کرد آموزش را به سطوح بالاتر حیطه شناختی برساند. این مطالعه با استناد به پژوهش‌های علی‌فرد و کامیاب (۱۳۹۲) و سهرابلو (۱۳۹۷)، نقش مثبت فناوری در این زمینه را تأیید می‌کند. Newman و همکاران (2013) و McLaughlin و همکاران (2005) نیز بر نقش مؤثر ابزار جدول مشخصات آزمون در روایی محتوای مناسب آزمون صحه گماردند. سؤال دوم پژوهش: آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های طراحی سؤالات بر مبنای سطوح بالای شناختی را برطرف کند؟ در راستای پاسخ به سؤال پژوهش، پژوهشگر با مطالعه مبانی نظری و ابزارهای موجود، یک الگوی ترکیبی از طبقه‌بندی هدف‌های رفتاری بر اساس سه نظام شناختی، فراشناختی و خود طراحی کرد. این الگو با استفاده از چرخه‌های پروتوتایپ و مصاحبه با متخصصین و

معلمان ارزیابی و بهبود یافت. همچنین با کمک هوش مصنوعی chatGPT-4، فعل‌های رفتاری و هدف‌های رفتاری مناسب برای هر سطح شناختی توسعه یافت. در نهایت، روایی طبقه فعل‌های تولیدشده با استفاده از چک‌لیست و نظر متخصصان سنجش و اندازه‌گیری تأیید شد. هدف از این پژوهش، کمک به معلمان برای سوق دادن اندازه‌گیری‌های خود به سمت سطوح بالای شناختی بود که با رویکردهای نوین طبقه‌بندی هدف‌های رفتاری همخوانی دارد. راهکاری که پژوهش‌هایی همچون علی‌فرد و کامیاب (۱۳۹۲)، سهرابلو (۱۳۹۷)، Patel و همکاران (2006) همسو است؛ آن‌ها نشان دادند که از نظر اعضای هیئت‌علمی پس از استفاده از ابزار بلوپرینت آزمون (۷۹/۶۷٪)، سؤالات به‌خوبی سازمان‌دهی شده و دانش موضوعی عمیق (۷۴/۸۰٪) را آزموده است.

سؤال سوم پژوهش: آیا ابزاری برای توسعه ارزیابی آموزشی با رویکرد نرم‌افزارمحور می‌تواند برخی از چالش‌های ارتقای توانمندی معلمان در طراحی سؤالات با کیفیت را برطرف کند؟ این سؤال با زیرمجموعه‌هایی از جمله مخاطبان هدف، شیوه دسترسی ایشان به نرم‌افزار و جنبه‌های حمایتی که نرم‌افزار بر ارزیابی دارد مورد بررسی قرار گرفت.

نرم‌افزار جدول مشخصات آزمون بر پایه وب اپ طراحی شد که به معلمان (به‌ویژه معلمان درس علوم مقطع ششم ابتدایی) کمک کند تا با استفاده از جدول مشخصات آزمون، سؤالات مناسب و همسو با اهداف یادگیری و محتوای درسی خود را طراحی کنند. این نرم‌افزار بر اساس طبقه‌بندی ترکیبی بر مبنای رویکردهای اندرسون و همکاران و مارزانو و کندال ساخته شده است و نظام‌های شناختی، فراشناختی و خود را در نظر گرفته است. همچنین این نرم‌افزار با فراهم آوردن امکان تخصیص نسبت‌ها برای محتوای موردنظر، یک نمونه معرف را به معلمان ارائه می‌دهد. در طول فرآیند طراحی سؤالات، معلمان با فعل‌های رفتاری، هدف‌های رفتاری، قالب‌های سؤال و نمونه‌های مثال آشنا می‌شوند. این نرم‌افزار با توجه به یافته‌های Patel و همکاران (2006) که نشان داد که استفاده از بلوپرینت آزمون منجر به بهبود کیفیت سؤالات و انطباق آن‌ها با اهداف یادگیری می‌شود، تلاش کرده است تا به توسعه ارزیابی در آموزش کمک کند. ایشان نشان دادند که استفاده از ابزاری تحت عنوان بلوپرینت آزمون، وزن‌دهی مناسب نمرات برای موضوعات مهم اختصاص می‌یابد (۹۰٪)، سؤالات همسو با اهداف یادگیری طراحی می‌شوند (۸۰٪) توزیع سؤالات با توجه به اهمیت بالینی آن‌ها خواهد بود (۱۰۰٪) و سبب به حداقل رساندن تغییرات درون

آزمونگرها در گزینش سؤالات می‌شود (۹۰٪)؛ بنابراین استفاده از چنین ابزاری را برای دست‌اندرکاران آموزش از جمله معلمان مناسب دانسته‌اند.

کیفیت مناسب نرم‌افزار از جمله موضوعاتی است که مورد توجه پژوهشگر بوده است؛ ویژگی‌هایی از جمله عملکرد مناسب نرم‌افزار و جلب توجه کاربران؛ این نرم‌افزار می‌بایست در جهت اهدافی که به منظور آن‌ها طراحی شده است پیش رفته و مشخصاتی که برای آن طراحی شده است را برآورده سازد. جهت توسعه این نرم‌افزار می‌بایست نیازهای کاربران که در راستای اهداف طراحی آن است را در نظر گرفت؛ همچنین ویژگی‌هایی که یک نرم‌افزار برای داشتن کیفیتی قائل قبول باید داشته باشد نیز می‌بایست مدنظر قرار می‌گرفت و در همین راستا یک معماری مناسب و طراحی تفصیلی می‌توانست راهگشای خوبی باشد. همچنین تلاش بر آن بود که ضعف‌هایی که گاهی می‌توانند متوجه یک نرم‌افزار باشند برطرف گردند؛ از جمله عدم انطباق ویژگی‌های نرم‌افزار با نیازهای کاربران هدف (مطابق با رسالت نرم‌افزار)؛ یا عدم آگاهی از امکانات روز جهت پیاده‌سازی نرم‌افزار؛ جهت پوشش ضعف احتمالی اول سعی شد با یک دیدگاه علمی و بهره‌مندی از نظر متخصصان و امکانات روز دنیا از جمله هوش مصنوعی و پوشش گسترده مبانی نظری این احتمال کمرنگ گردد؛ جهت رفع احتمالی ضعف دوم نیز تلاش شد تا با مصاحبه با متخصصان نرم‌افزار بهترین راهکار جهت دسترسی حداکثری کاربران مختلف اتخاذ گردد و همچنین ویژگی‌هایی که بهترین کیفیت را در تأمین ویژگی‌های نرم‌افزار فراهم می‌کنند در نظر گرفته شوند؛ بنابراین سعی بر آن بود تا نرم‌افزار با کیفیتی قابل قبول بر مبنای اصول علمی و عملی طراحی شود و همان‌طور که ذکر آن رفت با شناسایی نیازهای کاربران با یک معماری مناسب در جهت رفع نیازها طراحی شود.

### تعارض منافع

نویسندگان هیچ‌گونه تعارض منافی ندارند.

### منابع

بروس. بی، فری (۱۳۹۵). سنجش کلاسی نوین (ترجمه: اسماعیل سعدی پور رها عابدی فاطمه رسولی). ناشر: موسسه فرهنگی مدرسه برهان



درتاج، فریبرز، و کرد نوقابی رسول. (۱۳۹۶). نظریه‌های آموزش: الگوها راهبردها، روش‌ها، فنون، ناشر: دانشگاه علامه طباطبایی

سهرابلو، مصطفی. (۱۳۹۷). ارزشیابی تکوینی الکترونیکی درس مطالعات اجتماعی. رشد آموزش علوم اجتماعی، پیاپی ۷۸ (زمستان ۱۳۹۷)، ص ۴۸. [magiran.com/p1941078](http://magiran.com/p1941078)

سیف، علی‌اکبر. (۱۳۶۴). اندازه‌گیری و ارزشیابی پیشرفت تحصیلی. تهران: موسسه انتشارات آگاه

سیف، علی‌اکبر. (۱۳۹۵). اندازه‌گیری، سنجش و ارزشیابی آموزشی. تهران: دوران، چاپ ۴۴

سیف، علی‌اکبر. (۱۳۹۵). روانشناسی پرورشی نوین. تهران: دوران.

طوفان زاده مؤددهی، علی، علایی، ابراهیم، و رشیدی، حسن. (۱۳۹۸). رهیافتی کاربردی در معماری نرم‌افزار، تهران: سبزیان گستر (چاپ دوم)

علی فرد، فاطمه، و کامیاب، شمس‌الدین. (۱۳۹۲). تأثیر فناوری‌های نوین بر آموزش زبان

انگلیسی در سطوح بالای حیطه شناختی بلوم، اولین کنفرانس ملی آموزش زبان انگلیسی،

ادبیات و مترجمی، شیراز، <https://civilica.com/doc/221973>

کیامنش، علیرضا؛ و گنجی، کامران. (۱۳۸۳). سنجش و یادگیری در دبیرستان. تهران: رشد

میگر. رابرت. او. (۱۳۵۵). تدارک هدف‌های آموزشی (ترجمه بهرام زنگنه و محمدحسین شمشیری). تهران: دانشگاه آزاد ایران (تاریخ انتشار اثر به زبان اصلی ۱۹۶۱).

نوروزی، داریوش و رضوی، سید عباس. (۱۳۹۰). مبانی طراحی آموزشی. تهران: سمت

نیکسون، رابین. (۱۳۹۹). مرجع کامل طراحی وب‌سایت با javascript, Mysql, PHP

HTML, CSS, jQuery (ترجمه یونس جمشیدی شیخ‌آبادی). انتشارات دانشگاهی

کیان

## References

- Ali Fard, F., & Kamiab, Sh. (2013). The Impact of Modern Technologies on Teaching English at Higher Levels of Bloom's Cognitive Domain. *First National Conference on English Language Teaching, Literature, and Translation, Shiraz*, <https://civilica.com/doc/221973>. [In Persian]
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Anderson, L. W., & Krathwohl, D. R. Airasian, P. W., Cruikshank, K. A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock. M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Ausubel, D. P., Stager, M., & Gaite, A. J. H. (1968). Retroactive facilitation in meaningful verbal learning. *Journal of Educational Psychology*, 59(4), 250.

- Biggs, J. B., & Collis, K. F. (2014). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals: handbook I: cognitive domain (No. 373.19 C734t)*. New York, US: D. Mckay. New York: McGraw-Hill
- Bridge, P. D., Musial, J., Frank, R., Roe, T., & Sawilowsky, S. (2003). Measurement practices: methods for developing content-valid student examinations. *Medical teacher*, 25(4), 414-421.
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Bruce B. F. (2013). *Modern classroom assessment*. Sage publications. [In Persian]
- Dortaj, F., & Kord Noghabi, R. (2016). *Theories of Instruction*, Allame Tabataba' University, Tehran, Iran [In Persian]
- Eggen, P., & Kauchak, D. (2001). *Educational Psychology: Windows on Classrooms*. 8th. Pearson
- Eweda, G., Bukhary, Z. A., & Hamed, O. (2020). Quality assurance of test blueprinting. *Journal of Professional Nursing*, 36(3), 166-170.
- Fink, L. D. (2013). *Creating significant learning experiences: An integrated approach to designing college courses*. John Wiley & Sons.
- Flagg, B. N. (2013). *Formative evaluation for educational technologies*. Routledge.
- Furst, E. J. (1994). Bloom's Taxonomy: Philosophical and Educational Issues. In Anderson, L.
- Gagné, R. M., & Gagné, R. M. (1985). *Conditions of learning and theory of instruction*. Holt, Rinehart and Winston.
- Gill, J. S., & Sen, S. (2020). Blueprinting of summative theory assessment of undergraduate medical students in microbiology. *Medical journal armed forces india*, 76(2), 207-212.
- Kaur, M., Kurmi, N., Chauhan, S., Singhal, A., Sharma, S., & Chaudhary, S. (2021). Blueprinting for assessment in undergraduate medical physiology curriculum. *Indian Journal of Physiology and Pharmacology*, 65(1), 60-65.
- Kiamanesh, A; & Ganji, K. (2004). *Assessment and Learning in High School*, Roshd Press, Tehran, Iran [In Persian]
- Mager, R. F. (1984). *Preparing Instructional Objectives (2nd edn)*, Pitman Learning Inc. [In Persian]
- Marso, R. N., & Pigge, F. L. (1993). Teachers' testing knowledge, skills, and practices.
- Marso, R. N., & Pigge, F. L. (1993). *Training and Management Perceptions of Public School Testing Directors: Implications for Teacher Educators*.
- Marzano, R. J., & Kendall, J. S. (Eds.). (2007). *The new taxonomy of educational objectives*. Corwin Press.
- McKenney, S. E. (2003). Computer-based support for science education materials developers in Africa: Exploring potentials.
- McLaughlin, K., Coderre, S., Woloschuk, W., & Mandin, H. (2005). Does blueprint publication affect students' perception of validity of the evaluation process? *Advances in health sciences education*, 10(1), 15-22.
- McLaughlin, K., Lemaire, J., & Coderre, S. (2005). Creating a reliable and valid blueprint for the internal medicine clerkship evaluation. *Medical teacher*, 27(6), 544-547.
- McLaughlin, K., Lemaire, J., & Coderre, S. (2005). Creating a reliable and valid blueprint for the internal medicine clerkship evaluation. *Medical teacher*, 27(6), 544-547.

- McMillan, J. (2007). *Classroom assessment: Principles and practices for effective standards-based instruction* (4th Ed.). Boston: Allyn & Bacon.
- Newman, I., Lim, J., & Pineda, F. (2013). Content validity using a mixed methods approach: Its application and development through the use of a table of specifications methodology. *Journal of Mixed Methods Research*, 7(3), 243-260.
- Nitko, A. (2004). *Educational assessment of students* (4th ed.). Upper Saddle River, NJ: Pearson.
- Nixon, R. (2012). *Learning PHP, MySQL, JavaScript, and CSS: A Step-by-Step Guide to Creating Dynamic Websites*. O'Reilly Media [In Persian]
- Norouzi, D., & Razavi, S. A. (2011). *Fundamentals of Instructional Design*, SAMT Press, Tehran, Iran. [In Persian]
- Patel, T., Saurabh, M. K., & Patel, P. (2016). Perceptions of the use of blueprinting in a formative theory assessment in pharmacology education. *Sultan Qaboos University Medical Journal*, 16(4), e475.
- Seif, A. A. (1985). *Measurement and Evaluation of Educational achievement*, Agah Press, Tehran, Iran [In Persian]
- Seif, A. A. (2016). *Educational Measurement, Assessment, and Evaluation*, Doran Press, Tehran, Iran. [In Persian]
- Seif, A. A. (2016). *Modern Educational Psychology*, Doran Press, Tehran, Iran. [In Persian]
- Sohrabloo, M. (2018). Formative Electronic Assessment of Social Studies Course. *Roshd Journal of Social Science Education*, No. 78 (Winter 2018), p. 48. [magiran.com/p1941078](http://magiran.com/p1941078) [In Persian]
- Tofanzadeh Mozhdehi, A., Alaei, E., & Rashidi, H. (2019). *A Practical Approach to Software Architecture*, 2nd edition, Sabz Rayan Gostar Press, Tehran, Iran. [In Persian]
- Wiggins, G., Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Asc.