



Developing a Teacher Assessment Literacy Model: A Case Study of
Farhangian University

A. Ramazani¹, M. S. Ghoraiishi Khorasgani², M. Sharifi³

1. Associate Professor, Department of Educational Management, Farhangian University, Tehran, Iran.
2. Department of Educational Administration and Planning, Faculty of Education and Psychology, Alzahra University, Tehran, Iran.
3. MA student, Faculty of Education and Psychology, Shahid Beheshti University, Tehran, Iran.

Article Info

ABSTRACT

Article type: Research

Article history:

Received: 19/11/2024
Revised: 25/10/2025
Accepted: 30/06/2026
Published: 21/03/2026

Keywords:

Assessment Literacy
Delphi Technique
Farhangian University
Systematic Review
Teacher training

Background and Objective: Assessment literacy (AL), encompassing conceptual, applied, and socio-emotional dimensions, is crucial for teachers to conduct effective classroom assessments and make evidence-based educational decisions. Accordingly, this study aimed to develop a comprehensive model of teachers' AL tailored to contemporary educational contexts. **Method:** This applied study employed a sequential mixed-methods design. In the qualitative phase, initial indicators were extracted through a systematic literature review, and in the quantitative phase, a three-round Delphi method was conducted with 31 purposively selected educational evaluation experts from Farhangian University across seven provinces in Iran, including Tehran, Zanjan, Sanandaj, Khuzestan, Mashhad, Isfahan, and Kerman. The data collection instrument was a researcher-made Delphi questionnaire. Data were analyzed using Statistical Package for the Social Sciences (SPSS) and Kendall's coefficient of concordance to measure expert consensus. **Findings:** The qualitative phase identified sub-indices for three main components, which were refined through the Delphi process. The finalized model achieved high expert consensus and comprises three main components, including conceptual, applied, and socio-emotional knowledge, nine sub-indices, and 33 operational indicators. **Conclusions:** The findings highlight the need to systematically integrate the conceptual, practical, and socio-emotional dimensions of AL into both pre-service and in-service teacher education programs. This comprehensive framework provides a robust foundation for the professional redefinition of teachers' roles in contemporary education systems, highlighting the significance of practical application of AL in both pre-service and in-service training.

Authors:

1. m.ghoraishi@alzahra.ac.ir

Citation (APA): Ramazani. A., Ghoraiishi Khorasgani. M.S., Sharifi. M.. (2026). Developing a Teacher Assessment Literacy Model: A Case Study of Farhangian University. *Journal of Teacher's Professional Development*, 11(1), 129-151.

 <https://doi.org/10.48310/tpd.2026.17611.1778>



© Author(s)

Publisher: Farhangian University

EXTENDED ABSTRACT

Background and Objectives: Assessment literacy (AL) as a fundamental and professional competency for teachers, enables them to design and implement effective classroom assessments, reduce assessment-related anxiety, and make evidence-based educational decisions. Consequently, it plays a strategic role in improving the quality of student learning. Beyond technical measurement skills, AL encompasses three dimensions, including conceptual, applied, and socio-emotional. Despite the development of numerous AL frameworks worldwide, most existing models have been developed within Western educational contexts and paid limited attention to the structural, cultural, and policy-specific characteristics of the Iranian education system, including the Fundamental Reform Document of Education. Moreover, previous research has paid limited attention identifying operational and context-specific indicators of these three dimensions within a comprehensive and localized model. Therefore, this study aimed to fill this theoretical and practical gap by developing a comprehensive, contemporary model tailored to Iranian educational context to provide a valid basis for strengthening teachers' AL, especially in the context of Farhangian University.

Methods: This study was applied in terms of its purpose and employed a sequential mixed-methods design consisting of qualitative and quantitative phases. In the qualitative phase, a systematic review of the literature was conducted to identify the initial indicators, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Relevant studies published between 2015 and 2025 were retrieved from national databases (SID, Magiran, Noormags, and IranDoc) and international databases (Scopus, Web of Science, ERIC, and PubMed). After the initial screening of 487 articles and the application of the inclusion and exclusion criteria, 28 eligible studies were finally selected for analysis resulting in the identification of 45 initial indicators. In the quantitative phase, a three-round Delphi technique was employed to validate and refine the identified indicators and to achieve expert consensus. The participants consisted of Farhangian University professors specializing in educational evaluation from seven provinces of the country, including Tehran, Zanjan, Sanandaj, Khuzestan, Mashhad, Isfahan, and Kerman. Purposive sampling was employed based on predefined expertise criteria. A total of 57 experts participated in the first Delphi round, 42 in the second round, and 31 completed the final round, resulting in an attrition rate of 45.6%, which is consistent with those reported in Delphi studies. The data collection instrument was a researcher-made Delphi questionnaire, which was distributed semi-open in the first round and closed in the second and third rounds with statistical feedback. The data were analyzed using SPSS version 27, and Kendall's coefficient of concordance (Kendall's W) was employed to assess the level of expert consensus and the stability of responses across the Delphi rounds.

Findings: The results showed that the final model of teachers' AL, with strong expert consensus (significant increase in the Kendall coefficient and a significance level of less than 0.001 in the third round), consisted of three main components, 9 sub-indices, and 33 operational indicators. The first component was *conceptual knowledge*, which included the sub-indices of theoretical understanding of evaluation, content analysis and standards,

and psychological knowledge of evaluation, and emphasizes teachers' mastery of theoretical foundations, standard frameworks, and psychological factors affecting learning. The second component was *applied knowledge*, which includes the sub-indices of designing assessment tools, implementing the assessment process, and analyzing and feedbacking the results, and guides teachers to design diverse tools, implement fair and ethical assessment, and use data to improve teaching. The third component is "social-emotional knowledge," which focuses on the emotional, communicative, and ethical aspects of evaluation with the sub-indices of creating a supportive evaluation environment, interacting with stakeholders, and managing emotions in evaluation. In the Delphi process, indicators, such as *adapting evaluation to local needs*, *managing anxiety and resilience*, *enhancing a sense of justice*, and *teacher emotional self-awareness* were finalized with special emphasis from the experts, showing that AL in the Iranian context requires a look beyond the purely technical aspect.

Conclusion: The findings of this study confirmed that AL can no longer be viewed merely as a technical and measurement-related competence. Rather, it represents a complex, contextual, and socio-emotional-ethical practice that directly impacts students' learning experiences and psychological well-being. The integration of conceptual, applied, and socio-emotional dimensions in the presented framework provides a solid foundation for a professional redefinition of the teachers' role in contemporary education systems and facilitates the transition from traditional assessment to learning-centered and equitable one. By moving beyond the direct adoption of Western frameworks and incorporating the ecological realities of Iranian classrooms and the policy context of the Iranian education system, this localized model provides a valid and practical framework for curriculum developers, educational policymakers, and instructors and faculty members at Farhangian University. Accordingly, the three dimensions of this model, particularly the often-overlooked socio-emotional dimension, should be systematically integrated into both pre-service and in-service teacher education programs to strengthen teachers' ability to design supportive assessments, reduce classroom anxiety, and enhance the quality of student learning. Future research is encouraged to evaluate this model in authentic classroom settings and examine its impact on students' learning and socio-emotional outcomes.



تدوین الگوی سواد سنجشی معلمان (مورد مطالعه: دانشگاه فرهنگیان)

عباس رضانی^{ID}، مریم سادات قریشی خوراسگانی^{ID}، معصومه شریفی^{ID}

۱. استادیار گروه مدیریت آموزشی، دانشگاه فرهنگیان، تهران، ایران.
۲. گروه مدیریت و برنامه ریزی آموزشی، دانشکده علوم تربیتی و روان شناسی، دانشگاه الزهرا، تهران، ایران.
۳. دانشجوی کارشناسی ارشد مدیریت آموزشی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید بهشتی، تهران، ایران.

اطلاعات مقاله چکیده

نوع مقاله: پژوهشی

تاریخچه مقاله:

تاریخ دریافت: ۱۴۰۳/۰۸/۲۹
تاریخ بازنگری: ۱۴۰۴/۰۸/۰۳
تاریخ پذیرش: ۱۴۰۵/۰۴/۰۹
تاریخ انتشار: ۱۴۰۵/۰۱/۰۱

واژه‌های کلیدی:

سواد سنجشی،
مرور سیستماتیک،
تکنیک دلفی،
آموزش معلمان،
دانشگاه فرهنگیان

نویسندگان:

1. m.ghoraishi@alzahra.ac.ir
(نویسنده مسئول)

پیشینه و اهداف: سواد سنجشی به‌عنوان توانایی معلمان در انجام ارزیابی‌های مؤثر کلاسی و اتخاذ تصمیمات آموزشی مبتنی بر شواهد، نقشی محوری در بهبود کیفیت یادگیری دارد. این سواد سه بعد کلیدی را در بر می‌گیرد: مفهومی، کاربردی و اجتماعی-عاطفی. هدف پژوهش حاضر، تدوین الگویی روزآمد و جامع از سواد سنجشی معلمان است. **روش‌ها:** این پژوهش از نوع کاربردی، با ماهیت توصیفی و روش آمیخته متوالی انجام شد. جامعه‌ی پژوهش را اساتید دانشگاه فرهنگیان در حوزه‌ی ارزشیابی آموزشی تشکیل دادند. در دو مرحله داده‌ها گردآوری شدند: در بخش کیفی، از طریق مرور سیستماتیک، شاخص‌های اولیه استخراج گردید؛ در مرحله‌ی دوم، با استفاده از پرسشنامه‌ی دلفی، نظرات ۳۱ نفر از اساتید دانشگاه فرهنگیان به عنوان حجم نمونه و با روش نمونه‌گیری هدفمند در هفت استان (تهران، زنجان، سمنان، خوزستان، مشهد، اصفهان و کرمان) جمع‌آوری شد. برای تحلیل داده‌ها از نرم‌افزار SPSS27 و آزمون ضریب همخوانی کندال جهت سنجش سطح اجماع خبرگان استفاده شد. **یافته‌ها:** در بخش کیفی، زیرشاخص‌های سه‌گانه برای سه مولفه‌ی اصلی شناسایی و در مراحل دلفی مورد توافق خبرگان قرار گرفت. در نهایت، چارچوبی جامع برای الگوی سواد سنجشی معلمان در نتیجه اجرای سه مرحله روش دلفی طراحی شد که شامل سه مولفه اصلی با ۹ زیرشاخص و ۳۳ شاخص استخراج‌شده از روش دلفی است.

نتیجه‌گیری: یافته‌ها بر ضرورت به‌کارگیری عملی سواد سنجشی در برنامه‌های آموزش معلمان تأکید دارند و نشان می‌دهند که یکپارچه‌سازی ابعاد مفهومی، کاربردی و اجتماعی-عاطفی می‌تواند پایه‌ای برای بازتعریف حرفه‌ای نقش معلم در نظام‌های آموزشی معاصر باشد.

استناد به این مقاله (APA): رضانی، عباس؛ نظری؛ قریشی خوراسگانی، مریم سادات و شریفی، معصومه. (۱۴۰۵). تدوین الگوی سواد سنجشی معلمان (مورد مطالعه: دانشگاه فرهنگیان). *توسعه حرفه‌ای معلم*, ۱۱(۱). ۱۵۱-۱۲۹.

DOI: 10.48310/TPD.2026.17611.1778



© نویسندگان.

ناشر: دانشگاه فرهنگیان

مقدمه

از جمله مسائلی که در نظام آموزش و پرورش مطرح است، موضوع ارزشیابی و چگونگی به کارگیری آن به منظور افزایش سطح آموزش و یادگیری است. آموزش و پرورش به عنوان نهادی اجتماعی هرگز نمی‌تواند از ارزشیابی بی‌نیاز گردد، پس در وجود ضرورت ارزشیابی در آموزش تردیدی نیست (دوست، ۱۴۰۲). بنابراین در هر نظام آموزشی، معلمان برای تعیین رابطه‌ی بین توانایی‌ها و فعالیت‌های دانش‌آموزان و تغییراتی که در رفتار آنان حاصل شده و نیز به منظور پیش‌بینی و کنترل موقعیت‌های آموزشی ارزشیابی را انجام می‌دهند (ظهرابی و همکاران، ۱۴۰۰). این نوعی از دانش تربیتی است، دانش تربیتی مربوط به روش‌ها و راهبردهایی است که معلمان برای آموزش موضوع‌های مختلف به دانش‌آموزان خود استفاده می‌کنند (رضانی و همکاران، ۱۴۰۳)؛ زیرا با ارزشیابی اشکالات و نقاط قوت و ضعف دانش‌آموزان مشخص می‌شود و این باعث بهبود دانش آن‌ها می‌شود. ارزشیابی از دیرباز مکمل فرایند تدریس و یادگیری بوده است (جیمز و همکاران^۱، ۲۰۰۲) و آن را موتور محرک یادگیری می‌نامند (کووان^۲، ۱۹۹۸). فرایند سنجش و ارزشیابی به موازات توسعه‌ی حرفه‌ای معلمان، اهمیت روزافزونی یافته است، از این رو، سیاست‌گذاران و برنامه‌ریزان نسبت به ایجاد مؤلفه‌های جدید سنجش و ارزشیابی برای ارتقای آموزش و یادگیری و افزایش مسئولیت‌پذیری در نظام آموزشی، توجه ویژه‌ای دارند (مهماندوست و همکاران، ۱۳۹۷). معلمانی که بیشتر از گذشته، بر الزام بکارگیری داده‌های ارزشیابی در فرایند تصمیم‌گیری خود واقف هستند، در مسیر پاسخگویی به نیازهای دانش‌آموزان فشار قابل توجهی را متحمل می‌شوند (استیگینز^۳، ۲۰۱۷)؛ زیرا آن‌ها از اهمیت انجام ارزشیابی مطلع هستند. از آن‌جا که توانش‌های ارزشیابی دانش‌آموزان منجر به تصمیم‌گیری‌های دقیق و آگاهانه می‌شود، سواد ارزشیابی از این حیث در کیفیت آموزش تأثیرگذار است (استاجی، ۱۴۰۰). ارزشیابی به عنوان یکی از ارکان اصلی نظام آموزشی، نقش تعیین‌کننده‌ای در کیفیت تدریس معلمان و یادگیری دانش‌آموزان ایفا می‌کند. این فرآیند نه تنها در سطح کلاس درس، بلکه در مقیاس سیستم مدرسه نیز به عنوان ابزاری اساسی برای پایش و ارتقای کیفیت آموزش شناخته می‌شود. ارزشیابی پیشرفت تحصیلی دانش‌آموزان، علاوه بر اینکه بازخوردهای مداومی را برای بهبود برنامه‌های درسی فراهم می‌کند (یادگارزاده و سرمدی انصار، ۱۳۹۲)، به عنوان ابزاری ضروری برای تضمین کیفیت آموزش و اثربخشی روش‌های تدریس عمل می‌کند (مهدی‌نیا و رسولی، ۱۴۰۲). از این رو، ارزشیابی نه تنها به عنوان ابزاری برای اندازه‌گیری یادگیری دانش‌آموزان، بلکه به عنوان سازوکاری استراتژیک برای توسعه حرفه‌ای معلمان و بهبود سیستم آموزشی مطرح است. هدف اصلی از ارزشیابی، توصیف تأثیر عمل‌های آموزشی و برنامه‌های درسی است (خزائی، ۱۴۰۲). این فرآیند شامل مراحل ماندند طراحی دقیق ارزشیابی، توضیح دلایل آن به دانش‌آموزان، و به اشتراک‌گذاری نتایج با همکاران است

1. James et al
2. Cowan
3. Stiggins

(استیگینز، ۲۰۱۷). بروکهارت^۱ (۲۰۱۱) با ارائه هفت استاندارد برای صلاحیت معلمان در ارزشیابی آموزشی، تأکید کرد که معلمان باید در انتخاب روش‌های مناسب، اجرای ارزشیابی، تفسیر نتایج، و استفاده از آن‌ها در تصمیم‌گیری‌های آموزشی تبحر داشته باشند. این استانداردها نه تنها به مهارت‌های فنی معلم می‌پردازند، بلکه به جنبه‌های اخلاقی و قانونی ارزشیابی نیز توجه دارند.

در سال‌های اخیر، سواد سنجشی به‌عنوان یکی از شایستگی‌های کلیدی حرفه‌ای معلمان مورد توجه گسترده‌ی پژوهشگران آموزشی قرار گرفته است. در این راستا، چندین مدل جهانی برای تبیین و تبیین ابعاد این سواد ارائه شده است. پاپهام^۲ (۲۰۰۹)، چارچوبی چندبعدی پیشنهاد کرد که دانش معلمان را در حوزه‌های هدف‌های یادگیری، روش‌های سنجش، اعتبار و پایایی ابزارها، و تفسیر نتایج ارزشیابی در بر می‌گیرد. در گامی فراتر، دلوکا و کلینگر^۳ (۲۰۱۰)، مدلی جامع‌تر ارائه دادند که سواد ارزشیابی را در چهار سطح دانش، مهارت، باور و عمل تعریف می‌کند و بر نقش باورهای فراشناختی و اخلاقی معلم در تصمیم‌گیری‌های ارزشیابی تأکید ویژه‌ای دارد. با الهام از نظریه‌ی اجتماعی-ساخت‌گرایی، پاپهام (۲۰۲۰) تعریفی چندبعدی از سواد ارزشیابی ارائه کرده است که سه بعد اصلی را در بر می‌گیرد: بعد مفهومی (درک مبانی نظری و فلسفی ارزشیابی)، بعد کاربردی (طراحی، اجرا و تحلیل ابزارهای سنجش)، و بعد اجتماعی-عاطفی (توجه به پیامدهای روانی، اخلاقی و عادلانه‌ی ارزشیابی بر دانش‌آموزان). این سه‌گانه توسط پژوهشگران دیگری نیز تأیید و گسترش یافته است؛ به‌طوری‌که بروکهارت (۲۰۱۱) بیشتر بر جنبه‌های مفهومی و فنی سواد سنجشی تمرکز کرده، در حالی که ژو و براون^۴ (۲۰۱۶) به‌ویژه بر ابعاد اجتماعی، عاطفی و اخلاقی آن تأکید دارند و نشان می‌دهند که ارزشیابی تنها یک فرآیند فنی نیست، بلکه در تعامل پیچیده‌ای با زمینه‌ی فرهنگی و روابط کلاسی شکل می‌گیرد.

با این وجود، بیشتر این مدل‌ها در بافت‌های آموزشی غربی توسعه یافته‌اند و به‌ندرت به ویژگی‌های ساختاری، فرهنگی و حرفه‌ای نظام‌های آموزشی در کشورهای در حال توسعه از جمله ایران پرداخته‌اند. همچنین، بیشتر مطالعات انجام شده به صورت کمی انجام شده‌اند و شاخص‌های اصلی این سه بعد را مورد بررسی قرار نداده و استخراج نکرده‌اند. بنابراین، شکاف اصلی پژوهشی این است که الگویی جامع و زمینه‌مند از شاخص‌ها و زیرشاخص‌های اصلی سواد سنجشی معلمان که سه بعد مفهومی، کاربردی و اجتماعی-عاطفی را در یک چارچوب یکپارچه در نظام آموزشی ایران در نظر بگیرد، تاکنون تدوین نشده است. بنابراین پژوهش حاضر در صدد پاسخ به این سوالات است که شاخص‌ها و زیرشاخص‌های اصلی سواد سنجشی معلمان چیست؟ و الگوی جامع این سواد چگونه است؟

1. Brookhart
2. Popham
3. Deluca & Klinger
4. Xu & Brown

روش

پژوهش حاضر، از لحاظ هدف کاربردی و از لحاظ ماهیت توصیفی و از لحاظ روش گردآوری داده‌ها آمیخته (کیفی-کمی) است. در ابتدا با استفاده از مرور سیستماتیک شاخص‌ها و زیرشاخص‌های مفهومی، کاربردی و اجتماعی-عاطفی سواد سنجشی معلمان استخراج شده است. در مرحله بعد برای شناسایی زیرشاخص‌های کلیدی برای تدوین الگوی سواد سنجشی معلمان، پرسشنامه محقق‌ساخته‌ای با طیف پنج سطح لیکرت (۱ = کاملاً مخالفم، ۲ = مخالفم؛ ۳ = تا حدودی موافقم، ۴ = موافقم، ۵ = کاملاً موافقم) طراحی شده و در بین اعضای خبرگان توزیع گردیده و از آن‌ها خواسته شده تا با توجه دیدگاه خود پاسخ‌های مناسبی به هر کدام از زیرشاخص‌ها ارائه نمایند. پس از گردآوری و انتقال داده‌ها برای اجرای روش دلفی از آزمون کندال w به منظور دستیابی به توافق بین خبرگان استفاده شده است.

به دلیل ایجاد امکان اجماع و تعیین اولویت‌ها، نخست از روش تحقیق دلفی، استفاده شده است. این روش، به صورت راه‌حلی برای ساختاردهی یک فرایند ارتباط گروهی تعریف می‌شود؛ به طوری که این فرایند به گروهی از افراد، به عنوان یک کل، امکان حل یک مسئله پیچیده را می‌دهد. هدف اصلی روش دلفی دستیابی به قابل اطمینان‌ترین اجماع گروهی از نظرات خبرگان به واسطه یک سری از پرسشنامه‌های متمرکز همراه با بازخورد کنترل شده می‌باشد. با کسب اجماع گروهی از خبرگان به وسیله این فرایند، پژوهشگران می‌توانند مسائل را شناسایی نموده و اولویت‌بندی کنند و چارچوبی را برای تشخیص آن‌ها توسعه دهند.

جامعه: جامعه‌ی آماری پژوهش حاضر، شامل تمامی اساتید دانشگاه فرهنگیان است که در زمینه‌ی ارزشیابی چه از منظر آموزشی (مانند ارزشیابی یادگیری دانشجویان، طراحی ابزارهای سنجش، یا بازخورد آموزشی) و چه از منظر پژوهشی (همچون ارزیابی برنامه‌های درسی، روش‌های سنجش اثربخشی تدریس، یا توسعه‌ی مدل‌های ارزشیابی) فعالیت علمی یا عملی دارند.

نمونه: نمونه‌ی پژوهش حاضر را اساتید و خبرگان دانشگاه فرهنگیان تشکیل دادند که به روش نمونه‌گیری هدفمند (قضاوتی) و بر اساس معیارهای ورود از پیش تعیین‌شده (شامل: حداقل ۱۰ سال سابقه تدریس در حوزه‌ی ارزشیابی آموزشی، داشتن تألیفات یا طرح‌های پژوهشی مرتبط با سنجش و اندازه‌گیری، و شناخت عمیق از بافت دانشگاه فرهنگیان) انتخاب شدند. در گام نخست فهرستی از تمامی اساتید واجد شرایط در سطح دانشگاه فرهنگیان تهیه و غربالگری شد.

در گام اولیه، ۵۷ نفر به عنوان پنل خبرگان انتخاب شدند. دلیل انتخاب این تعداد، از یک سو نیاز به پوشش دادن تمامی ابعاد و خرده‌مقوله‌های متنوع نظام ارزشیابی بود تا هیچ بُعدی از دیدگاه یک گروه تخصصی جا نماند، و از سوی دیگر، پیش‌بینی نرخ ریزش طبیعی در طول اجرای سه‌گانه‌ی دلفی بود؛ چرا که مطالعات پیشین نشان می‌دهد در فرآیندهای دلفی چندمرحله‌ای، ریزش تا ۵۰ درصد نیز امری رایج و پذیرفته‌شده است. فرآیند اجرا به این صورت بود که در دور اول، پرسشنامه‌ی باز در اختیار هر ۵۷ نفر قرار گرفت و نظرات آن‌ها استخراج شد. در دور دوم، با توجه به ماهیت طولانی‌تر فرآیند و نیاز به وقت بیشتر برای مطالعه و رتبه‌بندی گویه‌ها، تعداد ۴۲ نفر از خبرگان مشارکت کردند

(ریزش ۱۵ نفری معادل ۲۶.۳ درصد). در دور سوم و نهایی، که مستلزم دستیابی به اجماع و بازبینی مجدد بود، تعداد ۳۱ نفر از خبرگان تا پایان همراه بودند (ریزش ۱۱ نفری از دور دوم معادل ۲۶.۲ درصد). بدین ترتیب، نرخ ریزش کلی از دور اول تا دور سوم، ۴۵.۶ درصد (معادل ۲۶ نفر) گزارش می‌شود. این میزان ریزش، با توجه به سه مرحله‌ای بودن فرآیند دلفی و بازه زمانی اجرای پژوهش، در محدوده استاندارد مطالعات دلفی قرار دارد. در نهایت، نظرات ۳۱ نفر از خبرگان که در هر سه دور مشارکت کامل داشتند و به اجماع نهایی رسیدند، مبنای تحلیل‌ها و استخراج مدل قرار گرفت. ابزار: ابزار گردآوری شاخص‌ها در پژوهش حاضر در دو مرحله طراحی و تدوین شد. در وهله‌ی اول، مرور سیستماتیک ادبیات پژوهش با هدف شناسایی جامع شاخص‌های ارزشیابی آموزشی انجام گرفت. این مرور بر اساس پروتکل پریزما صورت پذیرفت. پایگاه‌های داده مورد جستجو شامل پایگاه‌های بین‌المللی Scopus، Web of Science، ERIC و PubMed و پایگاه‌های داخلی SID، Noormags Magiran و ایران‌داک بود. بازه‌ی زمانی جستجو از سال ۲۰۱۵ تا ۲۰۲۵ (معادل ۱۳۹۴ تا ۱۴۰۴) در نظر گرفته شد تا مطالعات به‌روز و مرتبط با تحولات اخیر در حوزه‌ی ارزشیابی آموزشی پوشش داده شود. واژگان کلیدی پژوهش و استراتژی جستجو در جدول (۱) ارائه گردید.

جدول (۱) استراتژی جستجو مقالات در پایگاه‌های اطلاعاتی

استراتژی جستجو به فارسی	استراتژی جستجو به انگلیسی
«ارزشیابی آموزشی» یا «سنجش» یا «اندازه‌گیری» یا «ارزشیابی عملکرد»	"educational evaluation" OR "assessment" OR "measurement" OR "performance evaluation"
و	AND
«ارزشیابی یادگیری» یا «دانشگاه» یا «آموزش عالی» یا «دانشگاه فرهنگیان»	"learning assessment" OR "University" OR "Higher education" OR "teacher education"

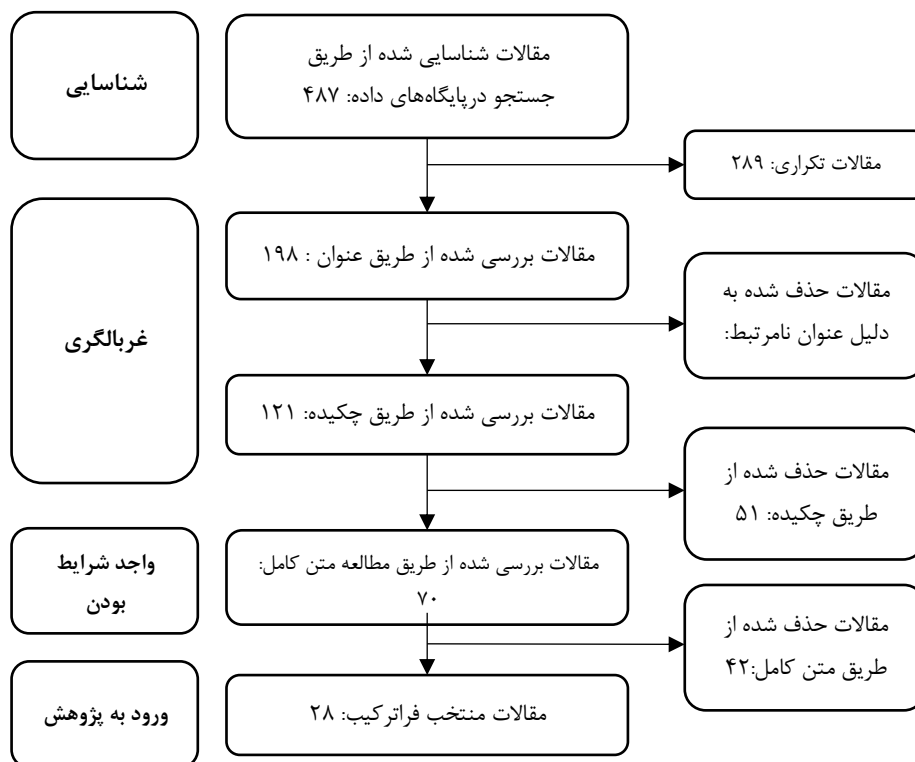
پس از این مرحله، به منظور انتخاب مقالات، معیارهای ورود و خروج تنظیم شد. این معیارها در جدول (۲) ارائه شده است.

جدول (۲) معیارهای ورود و خروج

ردیف	معیارهای ورود	معیارهای خروج
۱	مرتبط با واژگان کلیدی پژوهش	غیر مرتبط با حوزه مطالعاتی
۲	زبان فارسی و انگلیسی	همه زبان‌ها به غیر از فارسی و انگلیسی
۳	بازه‌ی زمانی (۲۰۱۵-۲۰۲۵)	خارج از بازه‌ی زمانی
۴	مطالعات مرتبط با سنجش و ارزشیابی	مطالعات مربوط به سایر حوزه‌ها
۶	امکان دسترسی تمام متن	عدم دسترسی تمام متن

با جستجوی واژگان کلیدی پژوهش مطابق استراتژی جستجو و با در نظر گرفتن معیارهای ورود ۴۸۷ مقاله شناسایی شد که ۲۸۹ مقاله در پایگاه‌ها به صورت تکراری نمایه شده بود. مقالات از جهت

عنوان مورد ارزیابی قرار گرفتند و ۷۷ مقاله به دلیل عنوان نامرتبط از فرایند کنار گذاشته شد. در گام بعدی چکیده ۱۲۱ مقاله باقی مانده مطالعه شد و ۵۱ مقاله از فرایند بررسی حذف شدند در نهایت با مطالعه متن کامل ۷۰ مقاله باقی مانده ۲۸ مقاله برای تحلیل و ترکیب نهایی انتخاب شدند. فرایند غربالگری مقالات در شکل (۱) ارائه گردید.

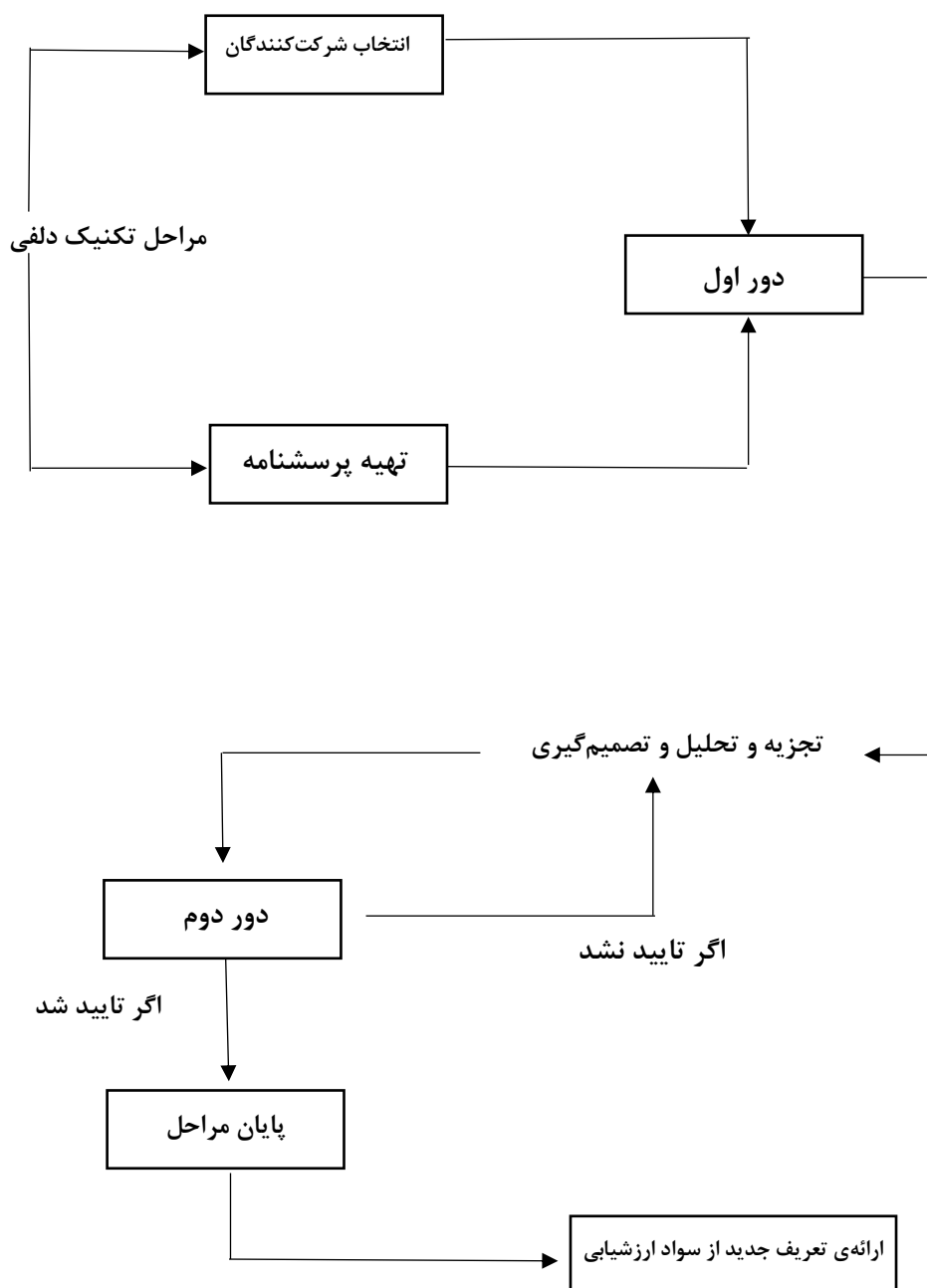


شکل (۱) فرایند گزینش مقالات (موهر، ۲۰۰۹)

در مرحله دوم، برای تکمیل، اعتبارسنجی و اجماع‌یابی درباره‌ی شاخص‌های استخراج‌شده، از روش دلفی استفاده شد. بدین منظور، پرسشنامه‌های محقق‌ساخته‌ای طراحی گردید که حاوی شاخص‌های اولیه حاصل از مرور سیستماتیک بود. این پرسشنامه‌ها در چند دوره بین نمونه‌ی توزیع شد. پس از هر دور، بازخوردهای خبرگان تحلیل و شاخص‌هایی که دارای سطح مشخصی از اجماع (مثلاً میانگین نظر بالاتر از آستانه‌ی تعیین‌شده یا انحراف معیار پایین) بودند، حفظ یا اصلاح گردیدند. این فرآیند تا حصول اجماع نسبی بین خبرگان ادامه یافت و در نهایت منجر به تدوین مجموعه‌ای معتبر و متناسب با زمینه‌ی پژوهش از شاخص‌های ارزشیابی شد.

روش تجزیه و تحلیل داده‌ها: سپس داده‌های جمع‌آوری‌شده از پرسشنامه‌های محقق‌ساخته با استفاده از نرم‌افزار SPSS 27 مورد تحلیل آماری قرار گرفت. به منظور سنجش درجه‌ی توافق بین خبرگان در روش دلفی، از ضریب همخوانی کندال استفاده شد. این ضریب، میزان اجماع را در

رتبه‌بندی شاخص‌های ارزشیابی توسط ۵۷ نفر از اساتید دانشگاه فرهنگیان نشان می‌دهد. برای آزمون معناداری ضریب کندال، از آماره‌ی کای اسکوئر استفاده گردید که با درجات آزادی محاسبه می‌شود. نتایج در سطح معناداری ۰/۵ مورد تفسیر قرار گرفتند. مراحل انجام پژوهش حاضر به شکل زیر است:



شکل (۲) مراحل انجام تکنیک دلفی

ابزار گردآوری داده‌ها، پرسشنامه‌ی محقق‌ساخته‌ی دلفی بود که در سه دور اجرا شد. در دور اول، پرسشنامه‌ی نیمه‌باز شامل ۴۵ گویه (مستخرج از مرور سیستماتیک) در طیف ۵‌گزینه‌ای لیکرت، به همراه بخش کیفی برای دریافت نظرات تکمیلی در اختیار خبرگان قرار گرفت. در دورهای دوم و سوم، پرسشنامه‌ها به صورت بسته و همراه با ارائه بازخورد آماری دور قبل بازطراحی شدند تا در نهایت ۳۳ گویه در ۹ زیرشاخص و ۳ مؤلفه نهایی گردد. روایی محتوایی این ابزار از دو مجرا تأیید شد: نخست، تبیین گویه‌ها بر اساس مرور سیستماتیک ۲۸ مقاله‌ی معتبر (روایی مبتنی بر منبع)؛ و دوم، قضاوت خبرگان در دور اول، به گونه‌ای که گویه‌های مبهم یا نامناسب (از نظر حداقل ۲۰ درصد خبرگان) اصلاح یا حذف شدند که عملاً معادل کیفی شاخص روایی محتوایی بود.

از آن جا که در روش دلفی گزارش پایایی کلاسیک رایج نیست، «ثبات پاسخ‌ها» در دورهای متوالی از طریق شاخص «همگرایی تدریجی» پایش و تضمین شد. این همگرایی از سه طریق عینی حاصل گردید: (۱) روند افزایشی ضریب همخوانی کندال در دورهای متوالی؛ (۲) کاهش انحراف معیار پاسخ‌ها به واسطه بازخورد کنترل‌شده آماری؛ و (۳) تثبیت ساختار گویه‌ها در دو دور پایانی که نشان‌دهنده‌ی اشباع نظری و رسیدن به اجماع نهایی بود.

در واقع مراحل انجام این بخش از پژوهش به شرح ذیل است:

- انتخاب نمونه‌ای از خبرگان
- تنظیم پرسشنامه دلفی و ارسال آن برای خبرگان
- گردآوری دیدگاه خبرگان و محاسبه میانگین دیدگاه آن‌ها
- تعیین شدت آستانه و حذف و اضافه شاخص‌ها
- تنظیم پرسشنامه دور دوم و افزودن میانگین نتایج دور نخست
- گردآوری پرسشنامه‌ها، تحلیل و ادامه دورها
- انتخاب سازوکاری پایان دورهای دلفی

تحلیل داده‌ها و ضریب میزان توافق از نرم‌افزار SPSS27 استفاده شد. با توجه به محدود بودن نمونه خبرگان و ماهیت این استعمال، تنها آمار توصیفی از میزان توافق بین شرکت‌کنندگان ارائه شده است. در نهایت برای پایان دورهای تکنیک دلفی، موافقت خبرگان با استفاده از ضریب همخوانی کندال (W) ارزیابی شد.

جدول (۳) ضریب همخوانی کندال (W)

تفسیر	مقدار W
عدم اتفاق نظر	۰
اندک	کمتر از ۰/۵
متوسط	بین ۰/۵ تا ۰/۷
خوب	بین ۰/۷ تا ۰/۹
خیلی زیاد	بین ۰/۹ تا ۱
اتفاق نظر کامل	۱

برای اطمینان از سازگاری، نظرات شرکت کنندگان توسط محققان از نظر موارد زیر مورد تجزیه و تحلیل قرار گرفت:

- شباهت‌ها و تفاوت‌های موجود در دیدگاه‌ها در مورد مؤلفه‌های اصلی سواد سنجشی
 - عناصر جدید پیشنهاد شده توسط کارشناسان
 - اصلاحات پیشنهادی برای اجزا
- بازنگری‌های پیشنهادی کارشناسان با گروه‌بندی آن‌ها بر اساس سه بعد اصلی سواد سنجشی (مفهومی، کاربردی و اجتماعی-عاطفی) از تعداد تجدیدنظرها کاست. پیشنهادات برای عناصر جدید نیز تجزیه و تحلیل، مقایسه، برچسب‌گذاری و به مناسب‌ترین بعد اختصاص داده شد. نهایتاً با استفاده از اصلاحات و بازنگری‌های انجام شده، تعاریف و در نتیجه ابعاد الگوی جدید استخراج گشت.

یافته‌ها

هدف از تحقیق حاضر تدوین الگوی سواد سنجشی معلمان از اساتید دانشگاه فرهنگیان استان‌های سنندج، تهران، زنجان، خوزستان، مشهد، اصفهان و کرمان در قالب سه بعد مفهومی، کاربردی و اجتماعی-عاطفی است. یافته‌های پژوهش حاضر برای تحقق این امر به تفکیک سوالات پژوهش ارائه خواهند شد.

سوال اول پژوهش: شاخص‌ها و زیرشاخص‌های اصلی سواد سنجشی معلمان چیست؟
در ابتدا با استفاده از مرور سیستماتیک شاخص‌ها و زیرشاخص‌های مفهومی، کاربردی و اجتماعی-عاطفی سواد سنجشی معلمان استخراج شده است. در مرحله اولیه استخراج شاخص‌های بعد منظر مفهومی سواد سنجشی معلمان شامل درک نظری ارزشیابی، تحلیل محتوا و استانداردها و دانش روان‌شناختی ارزشیابی است. در بعد کاربردی زیرشاخص‌های طراحی ابزارهای ارزشیابی، اجرای فرایند ارزشیابی و بازخورد نتایج ارزشیابی شناسایی شده است. همچنین بعد اجتماعی و عاطفی شامل زیرشاخص‌های ایجاد محیط ارزشیابی حمایتی، تعامل با ذی‌نفعان و مدیریت هیجانات در ارزشیابی است. همچنین در این مرحله ۴۵ زیرشاخص برای هر کدام از شاخص‌های اصلی شناسایی شده است. در جدول (۴) خلاصه‌ای از ابعاد سواد سنجشی، شاخص‌ها و زیرشاخص‌ها گزارش شده است.

جدول (۴) خلاصه‌ای از ابعاد سواد ارزشیابی، شاخص‌ها و زیرشاخص‌ها

مؤلفه اصلی	زیرمؤلفه	شاخص	منبع
د	درک نظری	تسلط بر مفاهیم ارزشیابی تکوینی و تراکمی	(استیگینز، ۲۰۱۷)
		آگاهی از اصول روانی و پایایی در طراحی ابزارهای ارزشیابی	(کرونباخ ^۱ ، ۱۹۷۱)
		تحلیل اسناد بالادستی ارزشیابی (مانند سند تحول بنیادین)	قاسم‌پور و همکاران (۱۳۹۸)
د	ارزشیابی	درک رابطه ارزشیابی با اهداف برنامه درسی	ویگینز و مک‌تایگ ^۲ ، (۲۰۰۵)
		شناخت چارچوب‌های نظری ارزشیابی (مانند طبقه‌بندی بلوم)	(بلوم ^۳ ، ۱۹۵۶)

1. Cronbach
2. Wiggins & McTighe
3. Bloom

ادامه جدول (۴) خلاصه‌ای از ابعاد سواد ارزشیابی، شاخص‌ها و زیرشاخص‌ها

منبع	شاخص	زیرمolfه	مولفه اصلی
(ویگینز و مک‌تایگ، ۲۰۰۵)	شناسایی اهداف یادگیری در محتوای درسی		
(پورتر و همکاران ^۱ ، ۲۰۱۱)	تسلط بر استانداردهای ملی و بین‌المللی ارزشیابی		
پیرکمالی و همکاران (۱۴۰۱)	تحلیل شکاف بین عملکرد دانش‌آموز و اهداف آموزشی	تحلیل محتوا و استانداردها	
(اندرسون و کراتول ^۲ ، ۲۰۰۱)	استفاده از چارچوب‌های نظری برای طراحی سوالات ارزشیابی		
(پورتر و همکاران، ۲۰۱۱)	تطبيق ارزشیابی با نیازهای محلی و بومی		
(ویگوتسکی ^۳ ، ۱۹۷۸)	آگاهی از نظریه‌های یادگیری مرتبط با ارزشیابی (مانند ویگوتسکی)		
(گاردنر ^۴ ، ۱۹۸۳)	درک تأثیر تفاوت‌های فردی بر نتایج ارزشیابی	دانش روان‌شناختی ارزشیابی	
(بندورا ^۵ ، ۱۹۹۷)	تفسیر داده‌های ارزشیابی با توجه به عوامل روان‌شناختی		
(هاتی ^۶ ، ۲۰۰۹)	ارائه بازخورد سازنده بر اساس اصول روان‌شناختی		
(دسی و رایان ^۷ ، ۲۰۰۰)	شناخت تأثیر انگیزه بر عملکرد در ارزشیابی		
(استیگینز، ۲۰۱۷)	طراحی آزمون‌های استاندارد (مانند چندگزینه‌ای یا تشریحی)	طراحی ابزارهای ارزشیابی	
(بروک‌هارت ^۸ ، ۲۰۱۳)	استفاده از ابزارهای متنوع ارزشیابی (مانند چک‌لیست، روبریک)		
(اندرسون و کراتول، ۲۰۰۱)	تنظیم سوالات متناسب با سطوح شناختی دانش‌آموزان		
(کرونباخ، ۱۹۷۱)	اطمینان از روایی و پایایی ابزارهای ارزشیابی		
(توماس ^۹ ، ۲۰۰۰)	طراحی ارزشیابی‌های مبتنی بر پروژه		
(مارزانو ^{۱۰} ، ۲۰۰۷)	مدیریت زمان و منابع در اجرای ارزشیابی		
(تومیلینسون ^{۱۱} ، ۲۰۰۱)	تنظیم ارزشیابی برای پاسخ به نیازهای متنوع دانش‌آموزان	اجرای فرآیند ارزشیابی	
(جانسون و جانسون ^{۱۲} ، ۱۹۹۹)	اجرای ارزشیابی‌های گروهی و فردی در کلاس		

جدول ۴

1. Porter et al
2. Anderson & Kratol
3. Vygotesky
4. Gardner
5. Bandura
6. Hattie
7. Deci & Ryan
8. Brookhurt
9. Tumas
10. Marzano
11. Tomilson
12. Jhomonson & Jhonson

ادامه جدول (۴) خلاصه‌ای از ابعاد سواد ارزشیابی، شاخص‌ها و زیرشاخص‌ها

مولفه اصلی	زیرمولفه	شاخص	منبع
		استفاده از فناوری در فرآیند ارزشیابی (مانند پلتفرم‌های آنلاین)	(بنت ^۱ ، ۲۰۱۱)
		رعایت اصول اخلاقی در اجرای ارزشیابی (مانند بی‌طرفی)	(پوفام ^۲ ، ۲۰۰۸)
		انطباق ارزشیابی با نیازهای دانش‌آموزان با نیازهای ویژه	(توملینسون، ۲۰۰۱)
		ارائه بازخورد شفاف و سازنده به دانش‌آموزان	(هاتی، ۲۰۰۹)
تحلیل و بازخورد نتایج		استفاده از نتایج ارزشیابی برای بهبود تدریس	بلک و ویلیام ^۳ ، (۱۹۹۸)
		مستندسازی و گزارش‌دهی نتایج ارزشیابی به ذی‌نفعان	(استیگینز، ۲۰۱۷)
		شناسایی الگوهای یادگیری از داده‌های ارزشیابی	بلک و ویلیام، (۱۹۹۸)
		ایجاد فضای امن و بدون استرس برای ارزشیابی	(راجرز ^۴ ، ۱۹۶۱)
		تشویق مشارکت فعال دانش‌آموزان در ارزشیابی	(کوهن ^۵ ، ۱۹۹۴)
ایجاد محیط ارزشیابی		تقویت اعتماد به نفس دانش‌آموزان در ارائه عملکرد	(بندورا، ۱۹۹۷)
		تقویت حس عدالت در فرآیند ارزشیابی	(رالز ^۶ ، ۱۹۷۱)
		مدیریت تعارضات و اضطراب مرتبط با ارزشیابی	(گولمن ^۷ ، ۱۹۹۵)
		استفاده از تشویق برای افزایش انگیزه در ارزشیابی	دسی و رایان، (۲۰۰۰)
دانش اجتماعی-عاطفی		برقراری ارتباط موثر با والدین در مورد نتایج ارزشیابی	(اپستین ^۸ ، ۲۰۰۱)
		جلب مشارکت دانش‌آموزان در خودارزیابی و همسال‌ارزیابی	(فالچیکوف ^۹ ، ۲۰۰۵)
		پاسخگویی به بازخوردهای ذی‌نفعان (مانند والدین، مدیران)	(اپستین، ۲۰۰۱)
		ایجاد شبکه‌های ارتباطی با جامعه برای ارزشیابی	(اپستین، ۲۰۰۱)
		کنترل هیجانات شخصی معلم در مواجهه با نتایج ضعیف	(گولمن، ۱۹۹۵)
مدیریت هیجانات در ارزشیابی		تقویت تاب‌آوری عاطفی دانش‌آموزان در برابر نتایج	(سلیگمن ^{۱۰} ، ۱۹۹۵)
		استفاده از تکنیک‌های انگیزشی برای بهبود عملکرد	دسی و رایان، (۲۰۰۰)
		تشخیص و مدیریت استرس دانش‌آموزان در ارزشیابی	(لازاروس ^{۱۱} ، ۱۹۹۱)
		تقویت خودآگاهی عاطفی معلم در فرآیند ارزشیابی	(گولمن، ۱۹۹۵)

1. Bennet
2. Pupham
3. Black & William
4. Rogers
5. Kohen
6. Rawls
7. Golman
8. Epestein
9. Falchikov
10. Siligman
11. Lazarus

سوال دوم پژوهش: الگوی جامع سنجش سواد سنجشی معلمان چگونه است؟

برای پاسخ به سوال دوم پژوهش حاضر، در مرحله بعد برای شناسایی زیرشاخص‌های کلیدی برای تدوین الگوی سواد سنجشی معلمان، پرسشنامه‌ای با طیف پنج سطح لیکرت (۱ = غیر ضروری ۲ = تاحدودی ضروری؛ ۳ = ضروری، ۴ = کاملاً ضروری) طراحی شده و در بین اعضای خبرگان توزیع گردیده و از آنها خواسته شده تا با توجه دیدگاه خود پاسخ‌های مناسبی به هر کدام از زیرشاخص‌ها ارائه نمایند. پس از گردآوری و انتقال داده‌ها برای اجرای روش دلفی از آزمون کندال W استفاده شده است. آزمون کندال یک آماره‌ی ناپارامتری است که برای سنجش درجه‌ی توافق (همخوانی) بین خبرگان در مورد رتبه‌بندی یک مجموعه از موارد به کار می‌رود (کندال و بابینگتون اسمیت^۱، ۱۹۳۹). در آزمون کندال برای استخراج زیرشاخص‌ها و توقف اجرای روش دلفی دو شرط در نظر گرفته شده است. نخست سطح معناداری کمتر از سطح خطای ۰/۰۵ باشد، دوم انحراف چارکی سوالات بالا نباشد. به طور کلی زیرشاخص‌هایی که میانگین آنها تفاوت معناداری با مقدار استاندارد ۳ نداشته و کمتر از این مقدار باشند، حذف می‌شوند. در جدول (۵) یافته‌های روش دلفی برای زیرشاخص‌های دانش مفهومی گزارش شده است. با توجه به اینکه در مرحله اول روش دلفی مقدار سطح معناداری ۰/۳۰۲ از سطح خطای ۰/۰۵ بیشتر است، بنابراین توافق نظری در بین خبرگان برای زیرشاخص‌های دانش مفهومی وجود ندارد. برای مرحله دوم روش دلفی مقدار سطح معناداری کمتر از سطح خطای ۰/۰۵ است که نشان‌دهنده برقراری توافق نظر خبرگان برای زیرشاخص‌های دانش مفهومی است، اما در این مرحله زیرشاخص‌های تحلیل اسناد بالادستی ارزشیابی، استفاده از چارچوب‌های نظری برای طراحی سوالات ارزشیابی، تفسیر داده‌های ارزشیابی با توجه به عوامل روان‌شناختی و تسلط بر استانداردهای ملی و بین‌المللی ارزشیابی به دلیل متفاوت بودن انحراف چارکی نسبت به شاخص‌های دیگر، از بین سوالات حذف شده‌اند. همچنین بیشتر پاسخ‌های خبرگان بر طیف ۳ (ضروری) و ۴ (کاملاً ضروری) اختصاص یافته است با این حال، در مرحله سوم توافق نظر خبرگان در حد قوی نیست.

جدول (۵) یافته‌های روش دلفی برای زیرشاخص‌های دانش مفهومی

مرحله اول	مرحله دوم	مرحله سوم	
۰/۰۱۹	۰/۰۵۲	۰/۱۱۴	کندال W
۱۶/۱۸۱	۴۳/۸۶۷	۷۵/۰۷۴	کای دو
۱۵	۱۵	۱۱	درجه آزادی
۰/۳۰۲	۰/۰۰۱	۰/۰۰۱	سطح معناداری

در جدول (۶) یافته‌های روش دلفی برای زیرشاخص‌های دانش کاربردی ارائه شده است. با توجه به این که در مرحله اول سطح معناداری ۰/۱۱۷ بیشتر از سطح خطای ۰/۰۵ است، بنابراین توافق نظری بین خبرگان برای زیرشاخص‌های دانش کاربردی وجود ندارد. در مرحله دوم مقدار سطح معناداری

1. Kendall & Babington Smith

کمتر از سطح خطای ۰/۰۵ است که نشان دهنده وجود توافق نظر ضعیف بین خبرگان است. با این حال در این مرحله زیرشاخص‌های استفاده از ابزارهای متنوع ارزشیابی، اطمینان از روایی و پایایی ابزارهای ارزشیابی، استفاده از فناوری در فرآیند ارزشیابی و تحلیل داده‌های ارزشیابی با روش‌های آماری به دلیل متفاوت بودن انحراف چارکی نسبت به شاخص‌های دیگر، از بین سوالات حذف شده‌اند. همچنین بیشتر پاسخ‌های خبرگان بر طیف ۳ (ضروری) و ۴ (کاملاً ضروری) اختصاص یافته است با این حال، در مرحله سوم توافق نظر خبرگان در حد قوی نیست.

جدول (۶) یافته‌های روش دلفی برای زیرشاخص‌های دانش کاربردی

مرحله سوم	مرحله دوم	مرحله اول	
۰/240	0/075	0/024	کندال W
158/359	68/808	21/644	کای دو
11	15	15	درجه آزادی
0/00۱	0/001	0/117	سطح معناداری

در جدول (۷) یافته‌های روش دلفی برای زیرشاخص‌های دانش اجتماعی و عاطفی ارائه شده است. با توجه به اینکه در مرحله اول سطح معناداری ۰/۰۶۹ بیشتر از سطح خطای ۰/۰۵ است، بنابراین توافق نظری بین خبرگان برای زیرشاخص‌های دانش اجتماعی و عاطفی وجود ندارد. در مرحله دوم مقدار سطح معناداری کمتر از سطح خطای ۰/۰۵ است که نشان دهنده وجود توافق نظر بین خبرگان است. با این حال زیرشاخص‌های فضای امن و بدون استرس برای ارزشیابی، همکاری با سایر معلمان برای هماهنگی در ارزشیابی‌ها، تقویت اعتماد به نفس دانش‌آموزان در ارائه عملکرد و پاسخگویی به بازخوردهای ذی‌نفعان به دلیل متفاوت بودن انحراف چارکی نسبت به شاخص‌های دیگر، از بین سوالات حذف شده‌اند. همچنین بیشتر پاسخ‌های خبرگان بر طیف ۳ (ضروری) و ۴ (کاملاً ضروری) اختصاص یافته با این حال، در مرحله سوم توافق نظر خبرگان در حد قوی نیست.

جدول (۷) یافته‌های روش دلفی برای زیرشاخص‌های دانش اجتماعی عاطفی

مرحله سوم	مرحله دوم	مرحله اول	
۰/۱۸۰	۰/۰۵۷	۰/۰۲۶	کندال W
۱۰۸/۰۰۱	۵۲/۳۲۵	۲۳/۷۷۷	کای دو
۱۰	۱۴	۱۴	درجه آزادی
۰/۰۰۱	۰/۰۰۱	۰/۰۶۹	سطح معناداری

جدول (۸) چارچوبی جامع برای الگوی سواد سنجشی معلمان در نتیجه اجرای سه مرحله روش دلفی را نشان می‌دهد که شامل سه مولفه اصلی (دانش مفهومی، دانش کاربردی، دانش اجتماعی-عاطفی) با ۹ زیرشاخص و ۳۵ شاخص استخراج‌شده از روش دلفی است. این چارچوب به‌منظور تقویت توانایی معلمان در ارزیابی یادگیری دانش‌آموزان، بهبود فرآیند تدریس، و ایجاد محیطی حمایتی برای

ارزشیابی طراحی شده است. تحلیل همسویی زیرشاخص‌ها و شاخص‌ها با پدیده محوری (سواد سنجشی معلمان) نشان می‌دهد که هر عنصر به‌طور مستقیم به اهداف آموزشی، به‌ویژه در زمینه نظام آموزش و پرورش ایران و چارچوب‌های بین‌المللی، مرتبط است. همسویی بر اساس ارتباط منطقی شاخص‌ها با زیرشاخص‌ها، انطباق با نظریه‌های علمی، و کاربرد عملی در طراحی ابزارهای ارزیابی و برنامه‌های تربیت معلم بررسی شده است. این تحلیل با تأکید بر سند تحول بنیادین آموزش و پرورش ایران و استانداردهای جهانی ارزشیابی، چارچوبی منسجم برای ارزیابی و ارتقای سواد سنجشی معلمان ارائه می‌دهد.

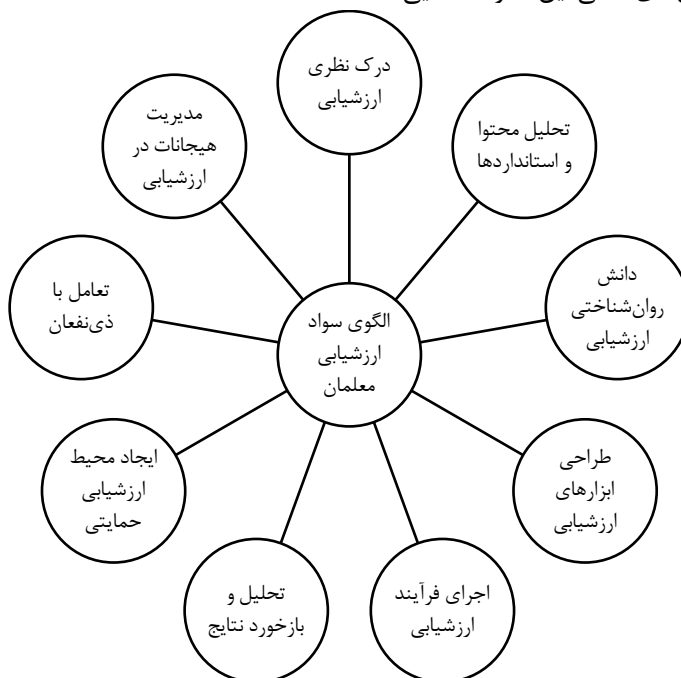
جدول (۸) زیرشاخص‌های نهایی استخراج شده از روش دلفی

میانگین رتبه	زیرشاخص	شاخص	مولفه اصلی	پدیده محوری
۶/۵۶	تسلط بر مفاهیم ارزشیابی تکوینی و تراکمی			
۶/۹۶	آگاهی از اصول روایی و پایایی در طراحی ابزارهای ارزشیابی	درک نظری		
۶/۹۷	شناخت چارچوب‌های نظری ارزشیابی	ارزشیابی		
۶/۵۴	درک رابطه ارزشیابی با اهداف برنامه درسی			
۶/۸۸	شناسایی اهداف یادگیری در محتوای درسی	تحلیل محتوا		
۶/۵۸	تطبیق ارزشیابی با نیازهای محلی و بومی	و استانداردها	دانش	
۳/۳۲	تحلیل شکاف بین عملکرد دانش‌آموز و اهداف آموزشی		مفهومی	
۶/۵۰	آگاهی از نظریه‌های یادگیری مرتبط با ارزشیابی (مانند ویگوتسکی)	دانش		
۶/۸۹	درک تأثیر تفاوت‌های فردی بر نتایج ارزشیابی	روان‌شناختی		
۷/۲۲	ارائه بازخورد سازنده بر اساس اصول روان‌شناختی	ارزشیابی		الگوی سواد ارزشیابی معلمان
۶/۷۸	شناخت تأثیر انگیزه بر عملکرد در ارزشیابی			
۷/۲۸	طراحی آزمون‌های استاندارد (مانند چندگزینه‌ای یا تشریحی)	طراحی ابزارهای ارزشیابی		
۷/۲۳	تنظیم سوالات متناسب با سطوح شناختی دانش‌آموزان			
۷/۲۳	طراحی ارزشیابی‌های مبتنی بر پروژه			
۲/۰۰	مدیریت زمان و منابع در اجرای ارزشیابی			
۷/۷۵	تنظیم ارزشیابی برای پاسخ به نیازهای متنوع دانش‌آموزان	اجرای فرآیند ارزشیابی		
۷/۶۰	اجرای ارزشیابی‌های گروهی و فردی در کلاس			دانش کاربردی
۷/۲۳	رعایت اصول اخلاقی در اجرای ارزشیابی (مانند بی‌طرفی)			
۶/۴۸	انطباق ارزشیابی با نیازهای دانش‌آموزان با نیازهای ویژه			
۶/۲۲	ارائه بازخورد شفاف و سازنده به دانش‌آموزان	تحلیل و بازخورد		
۶/۵۰	استفاده از نتایج ارزشیابی برای بهبود تدریس			
۶/۱۰	مستندسازی و گزارش‌دهی نتایج ارزشیابی به ذی‌نفعان	نتایج		
۶/۳۸	شناسایی الگوهای یادگیری از داده‌های ارزشیابی			

ادامه جدول (۸) زیرشاخص‌های نهایی استخراج شده از روش دلفی

میانگین رتبه	زیرشاخص	شاخص	مولفه اصلی	پدیده محوری
۵/۶۱	تشویق مشارکت فعال دانش‌آموزان در ارزشیابی	ایجاد محیط		
۵/۹۶	تقویت حس عدالت در فرآیند ارزشیابی	ارزشیابی		
۵/۹۹	مدیریت تعارضات و اضطراب مرتبط با ارزشیابی	حمایتی		
۶/۷۶	استفاده از تشویق برای افزایش انگیزه در ارزشیابی	تعامل با	دانش	
۶/۵۳	برقراری ارتباط موثر با والدین در مورد نتایج ارزشیابی	ذی‌نفعان	اجتماعی-	
۶/۵۸	ایجاد شبکه‌های ارتباطی با جامعه برای ارزشیابی		عاطفی	
۶/۶۴	کنترل هیجانات شخصی معلم در مواجهه با نتایج ضعیف	مدیریت		
۶/۳۶	تقویت تاب‌آوری عاطفی دانش‌آموزان در برابر نتایج	هیجانات در		
۶/۷۸	استفاده از تکنیک‌های انگیزشی برای بهبود عملکرد	ارزشیابی		
۶/۲۵	تشخیص و مدیریت استرس دانش‌آموزان در ارزشیابی			
۲/۵۴	تقویت خودآگاهی عاطفی معلم در فرآیند ارزشیابی			

در شکل (۳) الگوی سواد سنجشی معلمان گزارش شده است که درک نظری ارزشیابی، تحلیل محتوا و استانداردها، دانش روان‌شناختی ارزشیابی، طراحی ابزارهای ارزشیابی، اجرای فرآیند ارزشیابی، بازخورد نتایج، تحلیل و بازخورد نتایج، تعامل با ذی‌نفعان و مدیریت هیجانات در ارزشیابی، طراحی ابزارهای ارزشیابی، اجرای فرآیند ارزشیابی، بازخورد نتایج، تحلیل و بازخورد نتایج، تعامل با ذی‌نفعان و مدیریت هیجانات در ارزشیابی به عنوان شاخص‌های اصلی این الگو شناسایی شده‌اند.



شکل (۳) الگوی سواد سنجشی معلمان

بحث و نتیجه گیری

یافته‌های پژوهش حاضر نشان می‌دهد که سواد سنجشی معلمان، سازه‌ای چندبعدی و پویا متشکل از سه مؤلفه اصلی «دانش مفهومی»، «دانش کاربردی» و «دانش اجتماعی-عاطفی» است. با این حال، ارزش و نوآوری اصلی این مدل در فهرست کردن صرف شاخص‌ها نیست، بلکه در فرآیند بومی‌سازی و انطباق اکولوژیک آن با واقعیت‌های کلاس‌درس و سیاست‌های کلان نظام آموزشی ایران نهفته است. تحلیل دقیق نظرات خبرگان در فرآیند اعتباریابی نشان می‌دهد که چرا برخی از شاخص‌های اولیه استخراج‌شده از ادبیات جهانی، در مدل نهایی حذف یا تعدیل شدند. برای مثال، در مراحل اولیه پژوهش، شاخص «تسلط بر استانداردهای ملی و بین‌المللی ارزشیابی» در زیرشاخص تحلیل محتوا قرار داشت، اما خبرگان آن را حذف و بر «تحلیل اسناد بالادستی و انطباق با نیازهای محلی» تأکید ورزیدند. تحلیل این تغییر نشان می‌دهد که خبرگان، استانداردهای انتزاعی بین‌المللی را در سطح کلان و دور از دسترس عمل روزمره معلم در بافتار بومی دانسته‌اند. در مقابل، با تکیه بر یافته‌های قاسم‌پور و همکاران (۱۳۹۸)، تمرکز به سمت تحلیل اسناد بالادستی (سند تحول بنیادین) و شناسایی شکاف بین عملکرد دانش‌آموز و اهداف آموزشی بومی تغییر مسیر داد. این حذف هوشمندانه و جایگزینی آن با شاخص‌هایی نظیر تطبیق ارزشیابی با نیازهای محلی و بومی (پورتر و همکاران، ۲۰۱۱)، نوآوری پژوهش حاضر را در «عملیاتی‌سازی زمینه‌مند سواد سنجشی» آشکار می‌سازد و بررسی می‌کند که سواد سنجشی در نظام آموزشی ایران، بیش از آنکه نیازمند حفظ کردن تئوری‌های محض بین‌المللی باشد، نیازمند درک عمیق از سیاست‌های تربیتی بومی و چارچوب‌های روان‌شناختی منطبق با فرهنگ یادگیری دانش‌آموزان ایرانی است.

برای تبیین دقیق‌تر سهم و نوآوری این پژوهش، مقایسه تطبیقی مدل استخراج‌شده با مدل‌های مرجع جهانی، به‌ویژه مدل پاپهام (۲۰۲۰) و دلوکا (۲۰۱۶) ضروری است. در بررسی وجوه اشتراک، مدل حاضر در مؤلفه‌های «دانش مفهومی» و «دانش کاربردی» همگرایی بالایی با دیدگاه‌های پاپهام (۲۰۲۰) دارد. همان‌طور که پاپهام بر گذار از ارزشیابی سنتی به تکوینی، طراحی ابزارهای معتبر (استیگینز، ۲۰۱۷؛ اندرسون و کراتول، ۲۰۰۱) و استفاده از بازخورد برای بهبود یادگیری تأکید دارد، زیرشاخص‌های طراحی ابزار، مدیریت زمان و منابع (مارزانو، ۲۰۰۷) و تحلیل نتایج در مدل حاضر نیز این رویکرد تکنیکال و داده‌محور را بازتولید کرده‌اند. از سوی دیگر، در بعد توجه به بافتار و زمینه‌مند بودن ارزشیابی، مدل حاضر با دیدگاه‌های دلوکا (۲۰۱۶) مبنی بر «سواد سنجشی به‌عنوان یک عمل جامعه‌شناختی و زمینه‌مند» مشترک است؛ چرا که در مدل حاضر نیز معلمان تشویق می‌شوند تا محتوای درسی را با اهداف ارزشیابی هماهنگ کرده و ارزشیابی‌هایی طراحی کنند که با استانداردهای ملی و بافتار فرهنگی سازگار باشند (ویگینز و مک‌تایگ، ۲۰۰۵).

با این وجود، تمایز بنیادین و نوآوری منحصر به فرد پژوهش حاضر نسبت به مدل‌های پاپهام و دلوکا، در «شکل‌گیری و پررنگ شدن مؤلفه دانش اجتماعی-عاطفی» به‌عنوان رکنی مستقل و عملیاتی است. در حالی که پاپهام (۲۰۲۰) بیشتر بر جنبه‌های شناختی-کاربردی و فنی ارزشیابی متمرکز است و دلوکا (۲۰۱۶) اگرچه به بافتار اجتماعی اشاره می‌کند، اما آن را عمدتاً در قالب عاملیت معلم

صورت‌بندی می‌کند؛ مدل استخراج‌شده در پژوهش حاضر، «مدیریت هیجانات»، «تاب‌آوری»، «عدالت ارزشیابی» و «ایجاد محیط حمایتی» را به‌صورت شاخص‌های مجزا وارد ساحت سواد سنجشی کرده است. این رویکرد که ریشه در تلفیق نظریه‌های هوش عاطفی (گولمن، ۱۹۹۵)، روان‌شناسی انسان‌گرا (راجرز، ۱۹۶۱) و تأکیدات سند تحول بنیادین بر تربیت تمام‌ساحتی دارد، نشان می‌دهد که یک معلم باسواد در ایران، نه تنها باید ابزار بسازد و به بافتار آگاه باشد، بلکه باید بتواند استرس و اضطراب ارزشیابی را مدیریت کند (لازاروس، ۱۹۹۱)، تاب‌آوری دانش‌آموزان را تقویت نماید (سلیگمن، ۱۹۹۵) و با درک تفاوت‌های فردی (گاردنر، ۱۹۸۳) و تأثیر انگیزه (دسی و رایان، ۲۰۰۰)، تجربه‌ای عاطفی-اخلاقی و عادلانه (رالز، ۱۹۷۱) برای دانش‌آموز خلق کند.

علاوه بر این، مؤلفه دانش کاربردی و مفهومی در این مدل نیز با تکیه بر ابعاد روان‌شناختی غنی‌تر شده است. آگاهی از نظریه‌های یادگیری (ویگوتسکی، ۱۹۷۸) و شناخت طبقه‌بندی‌های شناختی (بلوم، ۱۹۵۶) به معلم کمک می‌کند تا ارزشیابی‌هایی مبتنی بر پروژه (توماس، ۲۰۰۰) و متنوع طراحی کند. در مرحله اجرا نیز، رعایت اصول اخلاقی (پوفام، ۲۰۰۸) و انطباق با نیازهای دانش‌آموزان با نیازهای ویژه (توملینسون، ۲۰۰۱) از طریق اجرای ارزشیابی‌های گروهی و فردی (جانسون و جانسون، ۱۹۹۹) محقق می‌شود. در نهایت، این چرخه با تعامل گسترده با ذی‌نفعان، از جمله ارتباط با والدین و همکاری با معلمان (اپستین، ۲۰۰۱؛ دوفور و همکاران، ۲۰۰۶) و ارائه بازخورد شفاف و سازنده (هاتی، ۲۰۰۹؛ بلک و ویلیام، ۱۹۹۸) تکمیل می‌گردد. در مجموع، این چارچوب با عبور از کپی‌برداری صرف از مدل‌های غربی و بومی‌سازی دقیق شاخص‌ها بر اساس نیازهای اکولوژیک و اسناد بالادستی ایران، ابزاری جامع، اعتباریابی‌شده و منحصر به فرد را ارائه می‌دهد که سواد سنجشی را از یک مفهوم صرفاً تکنیکال، به یک کنشگری روزمره، عاطفی، اخلاقی و یادگیری‌محور در کلاس‌های درس تبدیل می‌کند.

با وجود دستاوردهای نظری و کاربردی، پژوهش حاضر با محدودیت‌هایی نیز همراه بود که باید در تفسیر نتایج و تعمیم‌پذیری آن‌ها مدنظر قرار گیرد. نخست آنکه، با توجه به بافتار خاص نظام آموزشی ایران و تأکید ویژه بر اسناد بالادستی بومی (مانند سند تحول بنیادین)، تعمیم مدل استخراج‌شده به نظام‌های آموزشی با ساختارها، سیاست‌ها و فرهنگ‌های متفاوت نیازمند احتیاط و انجام پژوهش‌های تطبیقی است. بر اساس یافته‌ها و محدودیت مطرح‌شده، پیشنهاداتی برای پژوهش‌های آتی ارائه می‌شود تا مسیر توسعه این حوزه هموارتر گردد. در گام نخست، پیشنهاد می‌شود در مطالعات آینده، با طراحی و روایی‌سنجی ابزارهای کمی (پرسشنامه‌های استاندارد) بر اساس مدل سه‌مؤلفه‌ای حاضر (دانش مفهومی، کاربردی و اجتماعی-عاطفی)، وضعیت سواد سنجشی معلمان را در مقیاسی وسیع‌تر و در سطوح مختلف تحصیلی و مناطق جغرافیایی گوناگون مورد سنجش و مقایسه قرار دهند. در گام دوم، انجام پژوهش‌های مداخله‌ای و نیمه‌تجربی با هدف طراحی و اجرای دوره‌های آموزش ضمن خدمت مبتنی بر این مدل، می‌تواند تأثیر ارتقای این ابعاد (به‌ویژه بعد اجتماعی-عاطفی) که در مدل‌های پیشین مغفول مانده بود) را بر کیفیت واقعی ارزشیابی در کلاس درس، کاهش اضطراب ارزشیابی و پیامدهای یادگیری دانش‌آموزان بررسی کند.

مشارکت نویسندگان

نویسنده‌ی اول: جمع‌آوری و تحلیل داده‌ها، نظارت بر پروژه، ویراستاری علمی پژوهش، روش‌شناسی پژوهش

نویسنده‌ی دوم: جمع‌آوری داده‌ها، نگارش متن، روش‌شناسی پژوهش،

نویسنده‌ی سوم: جمع‌آوری داده‌ها و اصلاحات

تشکر و قدردانی

نویسندگان مقاله بر خود لازم می‌دانند تا از تمامی شرکت‌کنندگان پژوهش، نهایت تشکر و قدردانی را به عمل آورند.

تعارض منافع

هیچگونه تعارض منافی توسط نویسندگان، بیان نشده است.

منابع

استاجی، معصومه. (۱۴۰۰). بررسی سواد ارزشیابی معلمان در کلاس‌های آموزش زبان انگلیسی به عنوان یک زبان خارجی: تأثیر پیشینه‌ی تحصیلی، فصلنامه‌ی اندازه‌گیری تربیتی، ۱۲(۴۵)، ۷-۳۸.

پیرکمالی، محمد علی؛ مومنی مهموئی، حسین و مهدی پور، یوسف. (۱۴۰۱). الگوی سواد سنجش و ارزشیابی برنامه درسی معلمان دوره ابتدایی: فرا ترکیب براساس مدل اروین. *مطالعات روانشناسی تربیتی*، ۱۹(۴۸)، ۳۸-۱۵. doi: 10.22111/jeps.2023.7319

خرزانی، آرمین؛ موسوی‌زاده، مهناز؛ بقائی‌نیا، مرضیه؛ نوربان حصار، کبری؛ بهرامی، فاطمه. (۱۴۰۲). بررسی اهمیت ارزشیابی پیشرفت تحصیلی در مدارس، فصلنامه‌ی پژوهشنامه مطالعات راهبردی علوم انسانی و اسلامی، ۴(۵۵)، ۲۴۷-۲۵۸.

دوست، فریبا. (۱۴۰۲). بررسی نقش و تاثیر ارزشیابی کیفی-توصیفی بر رشد و بروز خلاقیت دانش‌آموزان، فصلنامه‌ی مطالعات روانشناسی و علوم تربیتی، ۵(۵۶)، ۴۶۵-۴۷۸.

رمضانی، عباس؛ شریفی، معصومه؛ نصر، صبا و مقدمی، پریرسا. (۱۴۰۳). توسعه‌ی دانش حرفه‌ای معلمان: ادغام محتوا و آموزش در تربیت معلمان. *نوآوری‌های آموزشی*، ۲۳(۲)، ۵۷-۸۲. doi: 10.22034/jei.2024.413877.2736

قاسم‌پور، عرفانه؛ دهقانی، مرضیه؛ تلخایی، محمود؛ درانی، کمال و صالحی، کیوان. (۱۳۹۸). واکاوی نظریه شناخت موقعیتی و نسبت آن با تربیت معلم. *اندیشه‌های نوین تربیتی*، ۱۵(۱)، ۱۶۱-۱۸۶. doi: 10.22051/jontoe.2019.20185.2202

مهدی‌نیا، معصومه؛ رسولی، عصمت. (۱۴۰۲). دیدگاه فلسفی در مورد ارزشیابی تحصیلی، *مجله‌ی پیشرفت‌های نوین در روانشناسی، علوم تربیتی و آموزش و پرورش*، ۶(۶۶)، ۲۰۹-۲۲۴.

مهماندوست، زهره؛ علوی، محمد؛ کیوان‌پناه، شیوا. (۱۳۹۷). بررسی سواد سنجش و ارزشیابی مدرسان دانشگاهی زبان انگلیسی در ایران، فصلنامه‌ی پژوهش‌های زبانشناختی در زبان‌های خارجی، ۸(۲)، ۳۳۷-۳۳۵.

یادگارزاده، غلامرضا؛ سرمدی انصار، حسن. (۱۳۹۲). وزن دهی و اولویت‌بندی عوامل و نشانگرهای ارزشیابی برنامه درسی علوم تجربی دوره ابتدایی. پژوهش‌های آموزش و یادگیری (دانشور رفتار)، ۲۰(۲)، ۳۹-۶۰. SID. <https://sid.ir/paper/480456/fa>

- Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.
- Bandura, A. (1997). Self-efficacy: The exercise of control. Freeman.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. David McKay.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.
- Brookhart, S. M. (2013). *How to give effective feedback to your students* (2nd ed.). ASCD.
- Cohen, E. G. (1994). *Designing groupwork: Strategies for the heterogeneous classroom*. Teachers College Press.
- Cowan, J. (1998). *On becoming an innovative university teacher: Reflection in action*. Buckingham, Philadelphia, PA: Society for Research into Higher education & Open University Press.
- Cronbach, L. J. (1971). *Test validation*. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443-507). American Council on Education.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.
- Epstein, J. L. (2001). *School, family, and community partnerships: Preparing educators and improving schools*. Westview Press.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Goleman, D. (1995). *Emotional intelligence*. Bantam Books.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning at Australian Universities*. Melbourne, Australia: Center for the Study of Higher Education.
- Johnson, D. W., & Johnson, R. T. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5th ed.). Allyn & Bacon.
- Kendall, M. G., & Babington Smith, B. (1939). *The Problem of m Rankings*. *The Annals of Mathematical Statistics*, 10(3), 275-287.

- Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.
- Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. ASCD.
- Popham, W. J. (2008). Assessment literacy for teachers: Faddish or fundamental? *Theory and Practice*, 48(1), 4-11.
- Popham, W. J. (2020). *Classroom assessment: What teachers need to know* (9th ed.). Pearson.
- Porter, A. C., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards and assessment. *Educational Researcher*, 40(4), 186-198
- Rawls, J. (1971). *A theory of justice*. Harvard University Press
- Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy*. Houghton Mifflin.
- Seligman, M. E. P. (1995). *The optimistic child*. Knopf.
- Stiggins, R. J. (2017). The perfect assessment system. Alexandria, VA: ASCD. Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-26.
- Thomas, J. W. (2000). A review of research on project-based learning. *Journal of Experiential Education*, 23(1), 1-16.
- Tomlinson, C. A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). ASCD.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). ASCD.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A recon capitalization. *Teacher and teaching education*, 58 (1), 149-161.