



© ۱۴۰۵ تمامی حقوق انتشار این مقاله متعلق به نویسنده است.

انتشار این مقاله به صورت دسترسی آزاد مطابق با گواهی CC BY-NC 4.0 صورت گرفته است.

بررسی کمی عملکرد برنامه‌های درسی دوره ابتدایی در نظام آموزشی افغانستان از دیدگاه معلمان و کارشناسان

شیوه استناددهی: مسعودی، مسعود،

عباسی، عفت، حسینی‌خواه، علی، و حاتمی، جواد. (۱۴۰۵). بررسی کمی عملکرد برنامه‌های درسی دوره ابتدایی در نظام آموزشی افغانستان از دیدگاه معلمان و کارشناسان. یادگیری، تربیت و آموزش مدارس در هزاره سوم، ۳(۲)، ۱-۲۹.

مشخصات نویسندگان:

۱. مسعود مسعودی: دانشجوی دکتری برنامه‌ریزی درسی، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران

۲. عفت عباسی*: دانشیار، گروه مطالعات برنامه درسی و سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران

۳. علی حسینی‌خواه: دانشیار گروه مطالعات برنامه درسی و سنجش و اندازه‌گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه خوارزمی، تهران، ایران

۴. جواد حاتمی: استاد گروه علوم تربیتی، دانشکده علوم انسانی، دانشگاه تربیت مدرس، تهران، ایران

*پست الکترونیکی نویسنده مسئول: e.abbasi@khu.ac.ir

چکیده

هدف پژوهش حاضر، بررسی کمی عملکرد برنامه‌های درسی دوره ابتدایی افغانستان و مقایسه ارزیابی معلمان، مدیران، کارشناسان و متخصصان از مؤلفه‌های طراحی و تدوین، ساخت و تولید، اجرا، ارزشیابی و راهکارهای اصلاحی بود. این پژوهش کاربردی با روش توصیفی-پیمایشی و رویکرد همبستگی-مقایسه‌ای در میان کارکنان آموزش ابتدایی شش ولایت کابل، بامیان، مزارشریف، قندهار، فراه و هرات انجام شد. با نمونه‌گیری تصادفی طبقه‌ای، داده‌های ۵۷۲ معلم و مدیر و ۱۵۳ کارشناس و متخصص، در مجموع ۷۲۵ نفر، گردآوری شد. ابزارها شامل دو پرسش‌نامه محقق‌ساخته ۵۰ و ۳۰ گویه‌ای بود که روایی محتوایی و همگرا، پایایی ترکیبی و ساختار عاملی آن‌ها تأیید شد. داده‌ها با آزمون‌های t تک‌نمونه‌ای، t مستقل، فریدمن و تحلیل واریانس یک‌راهه در SPSS ۲۶ و SmartPLS ۳ تحلیل شدند. آزمون t تک‌نمونه‌ای نشان داد عملکرد کلی برنامه درسی از دیدگاه کارشناسان، $t(152)=25.82$ ، و معلمان و مدیران، $t(418)=56.20$ ، به‌طور معنی‌داری بالاتر از میانگین نظری بود ($p<0.01$) و تمامی مؤلفه‌ها نیز تفاوت معنی‌داری با مقدار ملاک داشتند. آزمون فریدمن تفاوت معنی‌دار رتبه مؤلفه‌ها را در کارشناسان ($\chi^2=114.18$)، $p<0.01$ و معلمان و مدیران ($\chi^2(4)=311.81$)، $p<0.01$ تأیید کرد. مقایسه دو گروه نشان داد در طراحی و تدوین تفاوت معنی‌داری وجود نداشت، $t(570)=-0.95$ ، $p=0.34$ ، اما در ساخت، $t(570)=-3.54$ ، $p<0.01$ ، و پیشنهادها و راهکارها، $t(570)=-2.95$ ، $p<0.01$ ، تفاوت معنی‌دار بود. همچنین سن، سابقه خدمت و ولایت بر ارزیابی برخی مؤلفه‌ها اثر معنی‌دار داشتند، درحالی‌که برای جنسیت و سطح تحصیلات اثر معنی‌داری مشاهده نشد. عملکرد برنامه درسی دوره ابتدایی افغانستان در مجموع بالاتر از سطح متوسط ارزیابی شد؛ با این حال، ناهمگونی معنی‌دار میان مؤلفه‌ها و تفاوت دیدگاه ذی‌نفعان نشان می‌دهد کیفیت برنامه درسی در همه مراحل یکسان نیست. بنابراین، اصلاحات آینده باید به‌طور ویژه بر بهبود فرایند ساخت و تولید کتاب‌های درسی، تقویت اجرای برنامه، بازنگری سازوکارهای ارزشیابی و تبدیل پیشنهادهای اصلاحی ذی‌نفعان به اقدامات اجرایی متمرکز شود.

واژگان کلیدی: عملکرد برنامه درسی؛ دوره ابتدایی؛ نظام آموزشی افغانستان؛ ارزشیابی برنامه درسی؛ معلمان؛ کارشناسان.

Submit Date: 07 April 2026
 Revise Date: 07 June 2026
 Accept Date: 14 June 2026
 Publish Date: 22 June 2026



© 2026 the authors. This is an open access article under the terms of the [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/) License.



Quantitative Evaluation of the Performance of Primary School Curricula in the Afghan Education System from the Perspectives of Teachers and Experts

How to cite: Masoudi, M., Abbasi, E., Hosseini Khah, A., & Hatami, J. (2026). Quantitative Evaluation of the Performance of Primary School Curricula in the Afghan Education System from the Perspectives of Teachers and Experts. *Learning, Training, and Education in Schools in the Third Millennium*, 3(2), 1-29.

Authors' Information:

1. Masoud Masoudi: Ph.D. Student in Curriculum Planning, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran
2. Effat Abbasi*: Associate Professor, Department of Curriculum Studies and Assessment and Measurement, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran
3. Ali Hosseini Khah: Associate Professor, Department of Curriculum Studies and Assessment and Measurement, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran
4. Javad Hatami: Professor, Department of Educational Sciences, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran

Corresponding author's email: e.abbasi@khu.ac.ir

Abstract

This study aimed to quantitatively evaluate the performance of primary school curricula in Afghanistan and compare teachers', administrators', experts', and specialists' assessments of curriculum design and development, production, implementation, evaluation, and proposed reforms. This applied study employed a descriptive survey with a correlational-comparative approach among primary education personnel in Kabul, Bamyan, Mazar-e-Sharif, Kandahar, Farah, and Herat provinces. Using stratified random sampling, data were collected from 572 teachers and administrators and 153 experts and educational specialists, yielding a total sample of 725 participants. Two researcher-developed questionnaires containing 50 and 30 items were used. Their content and convergent validity, composite reliability, and factorial structure were confirmed. Data were analyzed using one-sample t tests, independent-samples t tests, Friedman tests, and one-way analysis of variance in SPSS 26 and SmartPLS 3. One-sample t tests indicated that overall curriculum performance was significantly above the theoretical midpoint among experts, $t(152)=25.82$, and teachers and administrators, $t(418)=56.20$, both $p<.01$; all individual curriculum components also differed significantly from the criterion value. Friedman tests demonstrated significant differences in component rankings among experts ($\chi^2=114.18$, $p<.001$) and teachers and administrators ($\chi^2(4)=311.81$, $p<.001$). Between-group comparisons showed no significant difference in curriculum design and development, $t(570)=-0.95$, $p=.34$, but significant differences emerged for production, $t(570)=-3.54$, $p<.01$, and proposed reforms, $t(570)=-2.95$, $p<.01$. Age, work experience, and province significantly affected assessments of selected components, whereas gender and educational level did not. Although the overall performance of Afghanistan's primary school curriculum was evaluated above the moderate level, significant variation across curriculum components and stakeholder groups indicates uneven effectiveness across the curriculum cycle. Reform initiatives should therefore prioritize textbook production, curriculum implementation, evaluation mechanisms, and the systematic translation of stakeholder recommendations into actionable educational policies.

Keywords: Curriculum performance; Primary education; Afghan education system; Curriculum evaluation; Teachers; Educational experts.

مقدمه

آموزش ابتدایی بنیادی‌ترین سطح نظام آموزش رسمی و یکی از تعیین‌کننده‌ترین مراحل شکل‌گیری توانایی‌های شناختی، اجتماعی، عاطفی و رفتاری کودکان است؛ زیرا بخش مهمی از مهارت‌های پایه‌ای خواندن، نوشتن، محاسبه، حل مسئله، برقراری ارتباط، خودتنظیمی و نگرش نسبت به یادگیری در همین دوره شکل می‌گیرد و کیفیت تجربه‌های آموزشی در این مقطع می‌تواند بر مسیر تحصیلی و اجتماعی فرد در سال‌های بعد اثر ماندگار داشته باشد. در این میان، برنامه درسی نه صرفاً مجموعه‌ای از کتاب‌ها و محتوای آموزشی، بلکه چارچوبی نظام‌مند برای تعیین اهداف، انتخاب و سازمان‌دهی محتوا، طراحی فرصت‌های یادگیری، هدایت فعالیت‌های تدریس و یادگیری و ارزیابی میزان تحقق اهداف آموزشی است. از این منظر، کیفیت برنامه درسی را باید در پیوند میان عناصر متعدد آن و نه صرفاً در سطح محتوای رسمی سنجید. برنامه‌ریزی درسی زمانی اثربخش است که اهداف آموزشی، محتوا، راهبردهای تدریس، منابع آموزشی، فعالیت‌های یادگیری و شیوه‌های ارزشیابی در یک نظام منسجم و هدفمند به یکدیگر متصل شوند و در عمل نیز قابلیت اجرا داشته باشند (Fathi Vajargah, 2023; Maleki, 2019). ادبیات کلاسیک و معاصر برنامه درسی نیز بر همین دیدگاه نظام‌مند تأکید دارد و برنامه درسی را پدیده‌ای چندبعدی می‌داند که در آن تصمیم‌های مربوط به اهداف، محتوا، سازمان‌دهی تجارب یادگیری، اجرا و ارزشیابی باید در ارتباط متقابل با یکدیگر بررسی شوند (Marsh & Willis, 2007; Ornstein & Hunkins, 2018). در سطح آموزش ابتدایی، اهمیت این هماهنگی دوچندان است؛ زیرا هرگونه نارسایی در طراحی، ساخت، اجرا یا ارزشیابی می‌تواند مستقیماً کیفیت فرصت‌های یادگیری کودکان را محدود سازد. همچنین، با توجه به تحولات فناورانه و ظهور شایستگی‌های جدید، برنامه‌های درسی امروز ناگزیرند علاوه بر سوادهای سنتی، آمادگی دانش‌آموزان برای مواجهه با محیط‌های دیجیتال و مهارت‌های نوظهور را نیز در نظر گیرند؛ به‌گونه‌ای که حتی موضوعاتی مانند سواد هوش مصنوعی و تعامل هدفمند با فناوری‌های مولد به‌عنوان بخشی از شایستگی‌های مورد انتظار نسل جدید مورد بحث قرار گرفته‌اند (Song, 2025). ازاین‌رو، ارزیابی عملکرد برنامه‌های درسی ابتدایی باید نه به‌عنوان یک اقدام اداری محدود، بلکه به‌عنوان بخشی از سازوکار تضمین کیفیت نظام آموزشی تلقی شود.

از منظر نظری، اثربخشی برنامه درسی تابع میزان انسجام میان برنامه قصدشده، برنامه اجراشده و نتایج واقعی یادگیری است. حتی یک برنامه رسمی با اهداف روشن و محتوای مناسب، در صورتی که در مرحله اجرا با محدودیت منابع، ضعف آمادگی حرفه‌ای معلمان، نبود امکانات آموزشی یا روش‌های نامتناسب ارزشیابی روبه‌رو شود، الزاماً به پیامدهای مطلوب منجر نخواهد شد. فولان تغییر آموزشی را فرایندی چندسطحی می‌داند که موفقیت آن در گرو هماهنگی میان سیاست، ساختار، منابع، ظرفیت حرفه‌ای و کنش بازیگران مدرسه‌ای است (Fullan, 2016). در همین راستا، پژوهش درباره هم‌راستایی برنامه قصدشده و برنامه اجراشده نشان داده است که برداشت و تفسیر معلمان از اهداف و محتوای رسمی، نقشی تعیین‌کننده در نحوه عملیاتی‌شدن برنامه در کلاس درس دارد (Phaeton & Stears, 2016). بنابراین، بررسی اسناد رسمی یا محتوای کتاب‌ها به‌تنهایی برای قضاوت درباره عملکرد واقعی یک برنامه درسی کافی نیست و ارزیابی باید دیدگاه مجریان مستقیم برنامه،

به‌ویژه معلمان و مدیران، را نیز در بر گیرد. افزون بر آن، کارشناسان و متخصصان آموزشی که در سطوح طراحی، تدوین، تولید و سیاست‌گذاری برنامه فعالیت می‌کنند، می‌توانند ابعادی را شناسایی کنند که از دید مجریان مدرسه‌ای کمتر قابل مشاهده است. چنین رویکردی با منطق ارزشیابی مبتنی بر استفاده همخوان است که ارزش یک ارزیابی را در تولید شواهد معنادار برای تصمیم‌گیری، اصلاح و بهبود برنامه می‌داند (Patton, 2008). در نتیجه، ترکیب دیدگاه ذی‌نفعان مختلف می‌تواند تصویر جامع‌تری از کیفیت و آسیب‌های برنامه درسی فراهم کند و نقاط توافق و اختلاف میان سطوح سیاست‌گذاری و اجرا را آشکار سازد.

یکی از مؤلفه‌های اساسی عملکرد برنامه درسی، کیفیت طراحی و تدوین آن است. طراحی مطلوب مستلزم آن است که اهداف آموزشی روشن، قابل دستیابی، متناسب با نیازهای رشدی دانش‌آموزان و هماهنگ با نیازهای جامعه باشند و محتوا نیز از نظر توالی، سطح دشواری، ارتباط مفهومی و تناسب با ویژگی‌های سنی یادگیرندگان سازمان‌دهی شود (Taghipour Zahir, 2022). برنامه‌ریزی آموزشی و درسی همچنین باید از منطق نیازسنجی، تعیین اولویت‌ها و تخصیص متناسب منابع پیروی کند تا اهداف رسمی به برنامه‌های قابل اجرا تبدیل شوند (Mohsenpour, 2021). در مرحله بعد، کیفیت ساخت و تولید مواد آموزشی، به‌ویژه کتاب‌های درسی، بر فرصت‌های واقعی یادگیری اثر می‌گذارد. پژوهش‌های تجربی نشان داده‌اند که کیفیت کتاب درسی می‌تواند با توانایی دانش‌آموزان در انجام تکالیف و حل مسائل آموزشی مرتبط باشد و ویژگی‌هایی مانند سازمان‌دهی محتوا، وضوح ارائه و تناسب تمرین‌ها نقش مهمی در کیفیت یادگیری ایفا می‌کنند (Sievert et al., 2021). این موضوع در نظام‌هایی که کتاب درسی اصلی‌ترین یا حتی تنها منبع سازمان‌یافته آموزش کلاس است اهمیت بیشتری پیدا می‌کند. در کنار محتوای رسمی، آنچه دانش‌آموزان از محیط مدرسه، تعاملات روزمره، ساختارهای اقتدار و الگوهای رفتاری می‌آموزند نیز بخشی از تجربه تربیتی آنان را تشکیل می‌دهد. بررسی برنامه درسی پنهان نشان داده است که فضای نهادی و پیام‌های غیررسمی آموزش می‌تواند بر رشد شخصیتی و اجتماعی یادگیرندگان اثرگذار باشند (Victor Ojorka et al., 2024). بنابراین، سنجش عملکرد برنامه درسی مستلزم توجه هم‌زمان به کیفیت رسمی طراحی و محتوا و نیز شرایط واقعی محیط یادگیری است.

اجرای برنامه درسی نیز حلقه واسط میان طراحی رسمی و پیامدهای آموزشی به شمار می‌رود. کیفیت اجرا به عواملی مانند شایستگی حرفه‌ای معلم، دسترسی به منابع و تجهیزات، اندازه کلاس، زمان آموزشی، انعطاف‌پذیری روش‌های تدریس و میزان پشتیبانی نهادی وابسته است. رویکردهای جدید به توسعه برنامه درسی نشان می‌دهند که ویژگی‌های روان‌شناختی و رفتاری دانش‌آموزان نیز باید در طراحی مداخلات آموزشی و برنامه‌های توسعه مدرسه مورد توجه قرار گیرند و برنامه درسی زمانی اثربخش‌تر است که نسبت به تفاوت‌های فردی و نیازهای واقعی دانش‌آموزان حساس باشد (Mallillin & Mallillin, 2024). در آموزش موضوعات سلامت و مهارت‌های زندگی نیز بر ضرورت پیوند مستقیم میان اهداف، تدریس و ارزشیابی تأکید شده و طراحی برنامه درسی بدون توجه به نحوه اجرا و سنجش یادگیری ناکافی دانسته می‌شود (Benes & Alperin, 2025). از سوی دیگر، توسعه حرفه‌ای معلمان یکی از پیش‌شرط‌های اصلی اجرای اثربخش هر برنامه اصلاحی است. شواهد جدید نشان می‌دهد آموزش ضمن خدمت همراه با مشارکت فعال معلمان در پژوهش عملی می‌تواند توان آنان را برای اجرای ارزشیابی تکوینی و استفاده از بازخوردهای آموزشی بهبود بخشد (Parmigiani et al., 2025). ارزشیابی تکوینی نیز زمانی کارکرد واقعی دارد که

بخشی از فرایند یاددهی-یادگیری باشد و اطلاعات حاصل از آن برای تنظیم تدریس و حمایت از پیشرفت دانش آموز به کار رود (Wiliam, 2011). بنابراین، ضعف در هر یک از حلقه‌های اجرا و ارزشیابی می‌تواند موجب ایجاد فاصله میان برنامه رسمی و تجربه واقعی دانش آموزان شود.

اهمیت سنجش عملکرد برنامه درسی در افغانستان با توجه به ویژگی‌های تاریخی و ساختاری نظام آموزشی این کشور برجسته‌تر است. آموزش و پرورش افغانستان طی دهه‌های گذشته تحت تأثیر جنگ، بی‌ثباتی سیاسی، تغییرات حکومت‌ها، محدودیت زیرساخت‌ها و تحولات مکرر در سیاست‌های آموزشی قرار داشته است. تاریخ آموزش در افغانستان نشان می‌دهد که توسعه آموزش رسمی در این کشور همواره با تغییرات اجتماعی و سیاسی و چالش‌های دسترسی، سازمان‌دهی و تداوم برنامه‌ها همراه بوده است (Andishmand, 2010). تجربه تاریخی نظام‌های آموزشی منطقه نیز نشان می‌دهد که دگرگونی‌های سیاسی و اجتماعی می‌توانند ساختار مدرسه، محتوای آموزشی و جهت‌گیری‌های تربیتی را به‌طور مستقیم تحت تأثیر قرار دهند (Aghazadeh, 2020). در افغانستان، اصلاحات آموزشی پس از ۲۰۰۱ موجب گسترش دسترسی به مدرسه و بازنگری‌هایی در نصاب تعلیمی شد، اما توسعه کمی الزاماً به رفع مشکلات کیفی منجر نشد. تحلیل برنامه درسی ملی مدارس ابتدایی افغانستان پس از ۱۱ سپتامبر نشان داده است که میان اهداف اعلام‌شده، جهت‌گیری‌های برنامه رسمی و برخی ابعاد محتوای واقعی آموزشی فاصله‌هایی وجود داشته است (Amiri, 2016). این وضعیت نشان می‌دهد که ارزیابی برنامه درسی افغانستان باید علاوه بر اسناد سیاستی، چگونگی ادراک و اجرای آن در سطح مدرسه و نیز کیفیت فرایندهای تولید و پشتیبانی را بررسی کند.

گزارش‌های بین‌المللی نیز ضرورت توجه به کیفیت آموزش ابتدایی افغانستان را تأیید می‌کنند. اطلاعات منتشرشده درباره وضعیت آموزش افغانستان حاکی از آن است که شمار زیادی از کودکان همچنان با محدودیت در دسترسی به آموزش روبه‌رو هستند و نابرابری‌های جغرافیایی، جنسیتی و اجتماعی، همراه با کمبود منابع و اختلال در فرصت‌های آموزشی، بر وضعیت یادگیری اثر گذاشته است (Unicef, 2025). گزارش‌های جهانی آموزش نیز نشان می‌دهند که دسترسی صرف به مدرسه برای تحقق عدالت آموزشی کافی نیست و کیفیت، تداوم، منابع، معلمان و فرصت‌های واقعی یادگیری باید هم‌زمان در کانون توجه قرار گیرند (Unesco, 2023). از منظر گسترده‌تر، نتایج ارزیابی‌های بین‌المللی پس از اختلال‌های آموزشی سال‌های اخیر نشان داده‌اند که نظام‌های آموزشی در بسیاری از کشورها با افت یادگیری و افزایش شکاف میان گروه‌های دانش‌آموزی مواجه شده‌اند و بازایی یادگیری نیازمند سیاست‌هایی است که کیفیت تدریس، پشتیبانی از مدارس و ارزیابی مستمر را یکپارچه سازد (Oecd, 2023). اگرچه افغانستان در بسیاری از ارزیابی‌های بین‌المللی استاندارد حضور محدود دارد، همین شواهد جهانی نشان می‌دهد که نظام‌های آموزشی آسیب‌پذیر بیش از دیگران به ابزارهای بومی و معتبر برای ارزیابی عملکرد برنامه درسی نیاز دارند تا تصمیم‌های اصلاحی به جای برداشت‌های کلی، بر شواهد کمی و مقایسه‌پذیر استوار شوند.

در چنین بستری، ارزشیابی چندمنبعی برنامه درسی اهمیت ویژه‌ای پیدا می‌کند. معلمان و مدیران به سبب تماس مستقیم با دانش‌آموزان و اجرای روزمره برنامه می‌توانند درباره قابلیت اجرا، کفایت منابع، دشواری محتوا، روش‌های تدریس و سازوکارهای ارزشیابی اطلاعاتی فراهم کنند که از طریق بررسی اسناد رسمی قابل دستیابی نیست. در مقابل، کارشناسان و متخصصان از منظر سیاستی و فنی قادرند کیفیت طراحی،

انسجام ساختار، استانداردهای تولید و ضرورت‌های اصلاح را ارزیابی کنند. مقایسه این دو دیدگاه نه تنها میزان توافق میان سطوح مختلف نظام آموزشی را نشان می‌دهد، بلکه می‌تواند نقاط کور هر گروه را نیز آشکار سازد. برای آنکه چنین مقایسه‌ای معتبر باشد، ابزار اندازه‌گیری باید از روایی و پایایی مناسب برخوردار باشد. در پژوهش‌های پیمایشی، روایی محتوا به‌ویژه هنگامی اهمیت دارد که ابزار محقق ساخته برای سنجش یک سازه چندبعدی تدوین می‌شود و شاخص‌هایی مانند CVI می‌توانند میزان توافق متخصصان درباره تناسب و نمایندگی گویه‌ها را نشان دهند (Polit & Beck, 2006). همچنین، ارزیابی ساختار اندازه‌گیری و سنجش روایی همگرا و پایایی سازه‌ها از الزامات تحلیل‌های کمی معتبر به شمار می‌رود و رویکردهای مبتنی بر مدل‌سازی معادلات ساختاری می‌توانند برای بررسی کیفیت مدل اندازه‌گیری و روابط میان سازه‌ها مورد استفاده قرار گیرند (Hair, 2014). توجه به این ملاحظات روش‌شناختی برای پژوهشی که قصد مقایسه ارزیابی چند گروه ذی‌نفع را دارد ضروری است، زیرا هرگونه ضعف در کیفیت ابزار می‌تواند تفاوت‌های مشاهده‌شده را به خطای اندازه‌گیری نسبت دهد، نه تفاوت واقعی در ادراک شرکت‌کنندگان.

با وجود اهمیت موضوع، مرور ادبیات نشان می‌دهد بخش قابل‌توجهی از مطالعات مربوط به آموزش افغانستان یا ماهیتی تاریخی، سیاستی و کیفی داشته‌اند یا بر بخش خاصی از برنامه درسی تمرکز کرده‌اند و شواهد کمی گسترده درباره عملکرد هم‌زمان مراحل مختلف برنامه درسی ابتدایی، به‌ویژه با مقایسه دیدگاه دو گروه اصلی ذی‌نفع، محدود است. از سوی دیگر، چارچوب‌های برنامه درسی تأکید می‌کنند که قضاوت درباره کیفیت یک برنامه باید از تمرکز بر یک مؤلفه منفرد فراتر رود و روابط میان طراحی، مواد آموزشی، اجرا و ارزشیابی را در نظر گیرد (Fathi Vajargah, 2023; Ornstein & Hunkins, 2018). رویکردهای اصلاح آموزشی نیز نشان می‌دهند که مداخلات موفق زمانی شکل می‌گیرند که نقاط ضعف هر مرحله شناسایی شود و میان سیاست‌گذاران، متخصصان و مجریان مدرسه‌ای نوعی درک مشترک از مسئله وجود داشته باشد (Fullan, 2016). همچنین، ارزشیابی زمانی بیشترین ارزش کاربردی را دارد که نتایج آن مستقیماً قابلیت استفاده در تصمیم‌گیری و اولویت‌بندی اصلاحات را داشته باشد (Patton, 2008). از این منظر، سنجش کمی عملکرد برنامه درسی ابتدایی افغانستان از دیدگاه معلمان، مدیران و کارشناسان می‌تواند هم وضعیت کلی را مشخص کند، هم مؤلفه‌های قوی‌تر و ضعیف‌تر را رتبه‌بندی نماید، هم تفاوت‌های ادراکی میان گروه‌های ذی‌نفع را آشکار سازد و هم نشان دهد که آیا ویژگی‌های حرفه‌ای و جمعیت‌شناختی شرکت‌کنندگان با نوع ارزیابی آنان از برنامه ارتباط دارد؛ چنین شواهدی می‌تواند مبنایی برای بازنگری هدفمند در محتوا، تولید کتاب‌های درسی، حمایت از اجرای برنامه، توسعه حرفه‌ای معلمان و اصلاح سازوکارهای ارزشیابی باشد.

بر این اساس، هدف پژوهش حاضر بررسی کمی عملکرد برنامه‌های درسی دوره ابتدایی در نظام آموزشی افغانستان از دیدگاه معلمان، مدیران، کارشناسان و متخصصان و مقایسه وضعیت مؤلفه‌های طراحی و تدوین، ساخت و تولید، اجرا، ارزشیابی و پیشنهادهای راهکارهای اصلاحی بود.

روش‌شناسی

پژوهش حاضر از نظر هدف، کاربردی و از نظر شیوه گردآوری داده‌ها، توصیفی از نوع پیمایشی با رویکرد همبستگی-مقایسه‌ای بود. جامعه آماری پژوهش را تمامی معلمان، مدیران، کارشناسان و متخصصان شاغل در بخش آموزش ابتدایی شش ولایت کابل، بامیان، مزارشریف، قندهار، فراه و هرات در سال تحصیلی ۱۴۰۳-۱۴۰۲ تشکیل دادند. بر اساس آمار اداره کل آموزش و پرورش کشور، جامعه معلمان این مناطق حدود ۱۹۸۰۰ نفر برآورد شده بود. برای دستیابی به نمونه‌ای که توزیع جغرافیایی جامعه هدف را به‌طور مناسب بازنمایی کند، از روش نمونه‌گیری تصادفی طبقه‌ای برحسب ولایت استفاده شد و حجم نمونه با توجه به حجم جامعه و بر پایه جدول مورگان تعیین گردید. در مجموع، ۶۵۰ پرسش‌نامه در میان معلمان و مدیران و ۱۷۰ پرسش‌نامه در میان کارشناسان و متخصصان آموزشی توزیع شد که به‌ترتیب ۵۹۰ و ۱۶۰ پرسش‌نامه بازگردانده شد؛ بنابراین، نرخ بازگشت پرسش‌نامه‌ها در دو گروه به‌ترتیب ۹۰.۸٪ و ۹۴.۱٪ بود. پس از بررسی اولیه، پرسش‌نامه‌هایی که بیش از ۱۰٪ گویه‌های آن‌ها بدون پاسخ مانده بود یا الگوی پاسخ‌دهی یکنواخت، یعنی انتخاب یک گزینه ثابت برای تمامی گویه‌ها، داشتند از مجموعه داده‌ها حذف شدند. در نهایت، داده‌های ۵۷۲ معلم و مدیر و ۱۵۳ کارشناس و متخصص آموزشی، در مجموع ۷۲۵ نفر، وارد تحلیل نهایی شد. در گروه معلمان و مدیران، ۷۱ نفر از کابل، ۱۰۷ نفر از بامیان، ۱۰۸ نفر از مزارشریف، ۱۰۶ نفر از قندهار، ۱۰۴ نفر از فراه و ۷۶ نفر از هرات مشارکت داشتند؛ در گروه کارشناسان و متخصصان نیز به‌ترتیب ۱۹، ۲۹، ۲۹، ۲۸، ۲۸ و ۲۰ نفر از این شش ولایت حضور داشتند. این توزیع تقریباً متناسب، از تمرکز بیش‌ازحد نمونه در یک منطقه خاص جلوگیری کرد و امکان مقایسه‌های میان‌ولایتی را فراهم ساخت. گردآوری داده‌ها در یک دوره چهارماهه و از طریق مراجعه حضوری به مدارس و ادارات معارف و نیز ارسال الکترونیکی پرسش‌نامه‌ها انجام شد. داده‌های گمشده پراکنده در گویه‌های منفرد که کمتر از ۲٪ کل داده‌ها را تشکیل می‌دادند، با روش جایگزینی میانگین سری در نرم‌افزار SPSS جایگزین شدند. با توجه به محدود بودن میزان داده‌های گمشده و پراکندگی آن‌ها، این روش برای جلوگیری از کاهش حجم نمونه در تحلیل‌های آماری به کار گرفته شد. در تمام مراحل پژوهش، اصول اخلاق پژوهش در علوم انسانی رعایت شد؛ مشارکت کاملاً داوطلبانه بود، هدف پژوهش و نحوه استفاده از داده‌ها پیش از تکمیل پرسش‌نامه برای شرکت‌کنندگان توضیح داده شد و حق انصراف از مشارکت در هر مرحله برای آنان محفوظ بود. برای تضمین محرمانگی، هیچ نام، شماره پرسنلی یا مشخصه قابل‌شناسایی دیگری ثبت نشد و نتایج صرفاً به‌صورت تجمیعی گزارش گردید. همچنین، با توجه به جایگاه سازمانی برخی کارشناسان در ساختار رسمی وزارت معارف، تأکید شد که پاسخ‌های فردی در اختیار مقامات مافوق قرار نخواهد گرفت تا احتمال تأثیر ملاحظات شغلی و سلسله‌مراتبی بر پاسخ‌ها به حداقل برسد. داده‌های کمی پژوهش با استفاده از دو پرسش‌نامه محقق‌ساخته گردآوری شد که بر اساس مرور نظام‌مند پیشینه نظری و تجربی پژوهش و مصاحبه‌های اکتشافی مقدماتی با متخصصان حوزه برنامه‌ریزی درسی و آموزش ابتدایی طراحی شده بودند. ابزار نخست، ویژه معلمان و مدیران بود و ۵۰ گویه بسته‌پاسخ را در پنج خرده‌مقیاس طراحی و تدوین، اجرا، ساخت و تولید، ارزشیابی، و پیشنهادها و راهکارها دربر می‌گرفت؛ هر خرده‌مقیاس شامل ۱۰ گویه بود، به‌گونه‌ای که گویه‌های ۱ تا ۱۰ به طراحی و تدوین، گویه‌های ۱۱ تا ۲۰ به اجرا، گویه‌های ۲۱ تا ۳۰ به

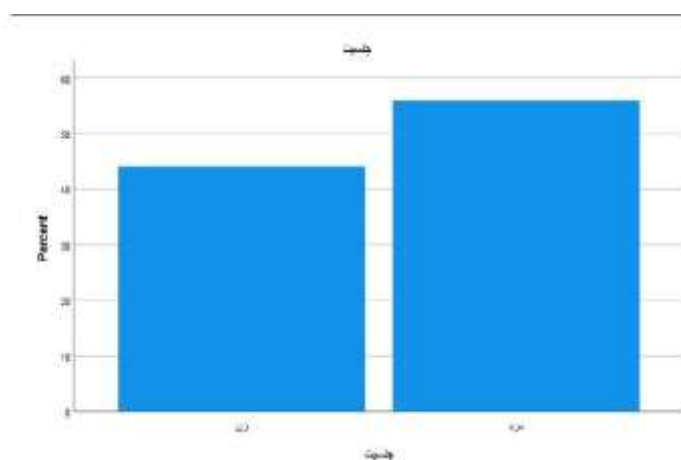
ساخت، گویه‌های ۳۱ تا ۴۰ به ارزشیابی و گویه‌های ۴۱ تا ۵۰ به پیشنهادها و راهکارها اختصاص داشت. ابزار دوم، ویژه کارشناسان و متخصصان آموزشی بود و شامل ۳۰ گویه در سه خرده‌مقیاس طراحی و تدوین، ساخت و تولید، و پیشنهادها و راهکارها بود که هر یک ۱۰ گویه داشتند؛ گویه‌های ۱ تا ۱۰ به طراحی و تدوین، گویه‌های ۱۱ تا ۲۰ به ساخت و گویه‌های ۲۱ تا ۳۰ به پیشنهادها و راهکارها اختصاص یافت. انتخاب دو ابزار مجزا بر مبنای تفاوت جایگاه حرفه‌ای، حوزه مسئولیت و میزان دسترسی اطلاعاتی دو گروه انجام شد؛ معلمان و مدیران به‌عنوان مجریان مستقیم برنامه درسی، امکان ارزیابی تمامی مراحل از طراحی و تولید تا اجرا و ارزشیابی کلاسی را داشتند، درحالی‌که کارشناسان و متخصصان بیشتر با سطوح کلان طراحی، تدوین و تولید محتوای آموزشی سروکار داشتند و به جزئیات روزمره اجرای کلاس و ارزشیابی دانش‌آموزان دسترسی یکسانی نداشتند؛ از این رو، خرده‌مقیاس‌های اجرا و ارزشیابی تنها در ابزار معلمان و مدیران منظور شد. پاسخ‌ها بر اساس طیف پنج‌درجه‌ای لیکرت از ۱ برای «کاملاً مخالفم» تا ۵ برای «کاملاً موافقم» نمره‌گذاری شدند و میانگین نظری ۳ به‌عنوان نقطه مرجع در نظر گرفته شد؛ بنابراین، میانگین‌های بالاتر از ۳ نشان‌دهنده ارزیابی مطلوب‌تر از وضعیت موجود و مقادیر پایین‌تر از ۳ نشان‌دهنده ارزیابی نامطلوب‌تر بودند. میانگین هر خرده‌مقیاس از مجموع نمرات گویه‌های مربوط و تقسیم آن بر تعداد گویه‌ها محاسبه شد. گویه‌ها به‌صورت مستقیم و عملیاتی تدوین شدند تا ابعاد مورد بررسی به‌طور مشخص سنجیده شوند؛ برای مثال، در حوزه طراحی و تدوین، وضوح اهداف آموزشی و تناسب محتوا با سن و توان یادگیری دانش‌آموزان، در حوزه ساخت، کیفیت چاپ و صحافی و هماهنگی تصاویر با محتوای متن، در حوزه اجرا، کفایت امکانات و تجهیزات کلاس برای اجرای برنامه، در حوزه ارزشیابی، توان روش‌های سنجش در اندازه‌گیری مهارت‌های واقعی دانش‌آموزان، و در حوزه پیشنهادها، ضرورت بازنگری محتوای کتاب‌های درسی و سایر راهکارهای اصلاحی مورد ارزیابی قرار گرفتند. پیش از اجرای اصلی، هر دو ابزار در یک مطالعه مقدماتی با مشارکت ۳۰ نفر از معلمان و کارشناسان خارج از نمونه اصلی بررسی شدند تا میزان وضوح عبارت‌ها، زمان موردنیاز برای تکمیل ابزار و وجود احتمالی گویه‌های مبهم یا دارای بار معنایی دوگانه مشخص شود. بر اساس بازخوردهای مرحله پایلوت، اصلاحات محدودی در عبارت‌بندی برخی گویه‌ها صورت گرفت، بدون آنکه ساختار مفهومی خرده‌مقیاس‌ها تغییر کند. روایی محتوایی ابزارها با نظر متخصصان، روایی همگرا بر اساس شاخص میانگین واریانس استخراج‌شده و پایایی آن‌ها با شاخص پایایی ترکیبی بررسی شد؛ همچنین ساختار عاملی ابزارها با تحلیل عاملی تأییدی مرتبه دوم ارزیابی شد و برازش مطلوب مدل‌های اندازه‌گیری مورد تأیید قرار گرفت. بدین ترتیب، نسخه نهایی پرسش‌نامه‌ها پس از طی مراحل بررسی مقدماتی روایی، پایایی و ساختار عاملی برای گردآوری داده‌های اصلی به کار رفت.

پس از پالایش، کدگذاری و ورود داده‌ها، تحلیل‌های آماری با استفاده از نرم‌افزار SPSS نسخه ۲۶ انجام شد و برای بررسی ساختار اندازه‌گیری و تحلیل عاملی تأییدی ابزارها از SmartPLS نسخه ۳ بهره گرفته شد. پیش از اجرای آزمون‌های استنباطی پارامتریک، مفروضه‌های موردنیاز بررسی شدند. نرمال بودن توزیع متغیرها با آزمون کولموگروف-اسمیرنوف و بررسی نمودارهای هیستوگرام و Q-Q ارزیابی شد و استقلال مشاهدات با توجه به طرح نمونه‌گیری تصادفی در نظر گرفته شد. همگنی واریانس‌ها در مقایسه‌های بین‌گروهی با آزمون لوین بررسی گردید و در مواردی که فرض برابری واریانس‌ها برقرار نبود، از نسخه اصلاح‌شده آزمون t برای واریانس‌های نابرابر بر اساس روش ولش استفاده شد. برای پاسخ به پرسش‌های اصلی پژوهش، ابتدا آزمون t تک‌نمونه‌ای به کار رفت تا میانگین عملکرد کلی برنامه درسی و میانگین هر یک از مؤلفه‌های

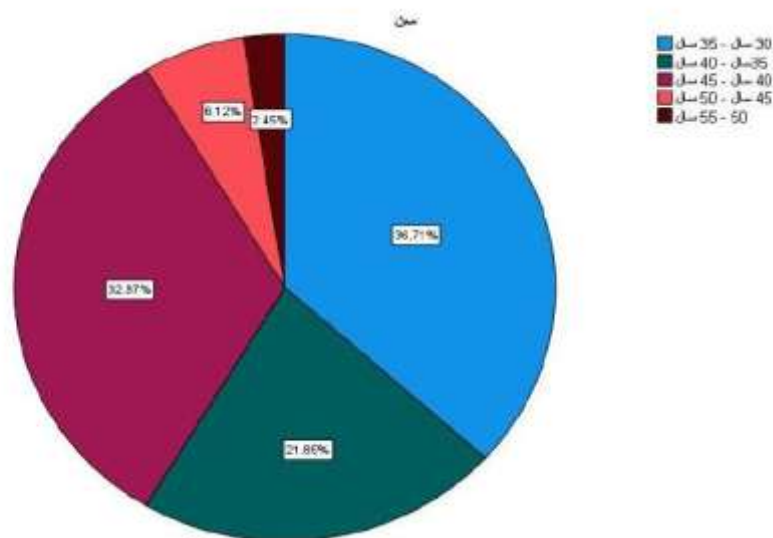
آن با میانگین نظری ۳ مقایسه شود و مشخص گردد ارزیابی شرکت‌کنندگان به‌طور معنی‌داری بالاتر یا پایین‌تر از سطح متوسط قرار دارد. سپس، برای مقایسه هم‌زمان و رتبه‌بندی مؤلفه‌ها در درون هر گروه از آزمون ناپارامتریک فریدمن استفاده شد؛ این آزمون با توجه به ماهیت رتبه‌ای مقایسه چند مؤلفه وابسته و عدم نیاز آن به مفروضه نرمال بودن برای این بخش از تحلیل مناسب تشخیص داده شد. برای مقایسه دیدگاه معلمان و مدیران با کارشناسان و متخصصان درباره سه مؤلفه مشترک شامل طراحی و تدوین، ساخت و تولید، و پیشنهادها و راهکارها، آزمون t برای گروه‌های مستقل به کار گرفته شد. همچنین، اثر متغیرهای جمعیت‌شناختی و حرفه‌ای چندسطحی، شامل سن، سابقه خدمت، ولایت محل خدمت و سطح تحصیلات، بر ارزیابی مؤلفه‌های برنامه درسی با استفاده از تحلیل واریانس یک‌راهه بررسی شد و در مواردی که تفاوت کلی معنادار بود، آزمون تعقیبی شفه برای تعیین محل دقیق تفاوت‌ها مورد استفاده قرار گرفت. اثر جنسیت نیز در قالب مقایسه‌های مناسب بین‌گروهی بررسی شد. تمامی آزمون‌های آماری دوطرفه در سطح اطمینان ۹۵٪ انجام شدند و مقدار ۰.۰۵ به‌عنوان سطح معناداری آماری در نظر گرفته شد. در کنار آزمون‌های استنباطی، شاخص‌های توصیفی از جمله میانگین، انحراف استاندارد، خطای استاندارد میانگین، کمینه و بیشینه نیز برای توصیف توزیع نمرات و فراهم کردن زمینه تفسیر نتایج محاسبه شدند.

یافته‌ها

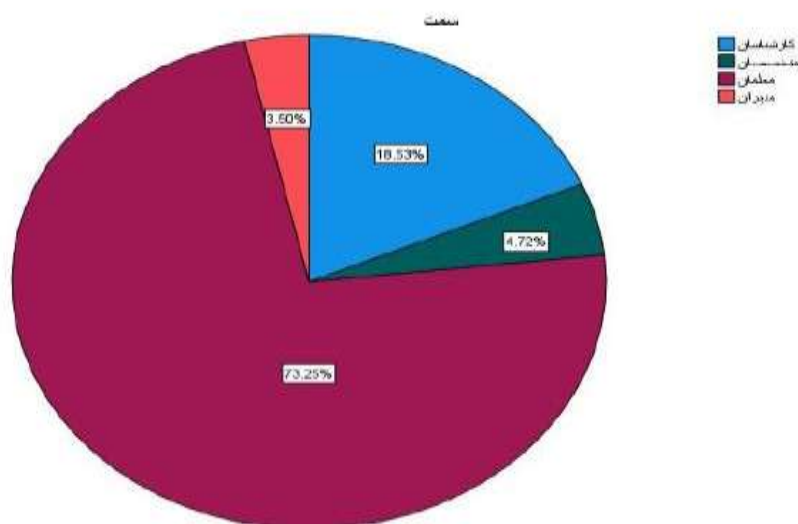
یافته‌های جمعیت‌شناختی نمونه مورد مطالعه نشان می‌دهد که از مجموع ۵۷۲ شرکت‌کننده در پژوهش، اکثریت مرد ($n = ۳۲۰$, ۵۵.۹٪) و بقیه زن ($n = ۲۵۲$, ۴۴.۱٪) بودند (نمودار ۱). بیشترین فراوانی سنی مربوط به گروه سنی ۳۰ تا ۳۵ سال ($n = ۲۱۰$, ۳۶.۷٪) و کمترین فراوانی مربوط به گروه سنی ۵۰ تا ۵۵ سال ($n = ۱۴$, ۲.۴٪) بود (نمودار ۲). از نظر سطح تحصیلات، بیشتر شرکت‌کنندگان دارای مدرک کارشناسی ($n = ۳۲۹$, ۵۷.۵٪) بودند و پس از آنان، دارندگان مدرک دیپلم و کاردانی ($n = ۲۰۸$, ۳۶.۴٪)، کارشناسی ارشد ($n = ۳۴$, ۵.۹٪) و دکتری ($n = ۱$, ۰.۲٪) قرار داشتند. به لحاظ سمت شغلی، معلمان ($n = ۴۱۹$, ۷۳.۳٪) بیشترین سهم را داشتند و پس از آنان کارشناسان ($n = ۱۰۶$, ۱۸.۵٪)، متخصصان ($n = ۲۷$, ۴.۷٪) و مدیران ($n = ۲۰$, ۳.۵٪) قرار داشتند (نمودار ۳).



نمودار ۱. فراوانی جنسیت نمونه آماری (معلمان و مدیران)

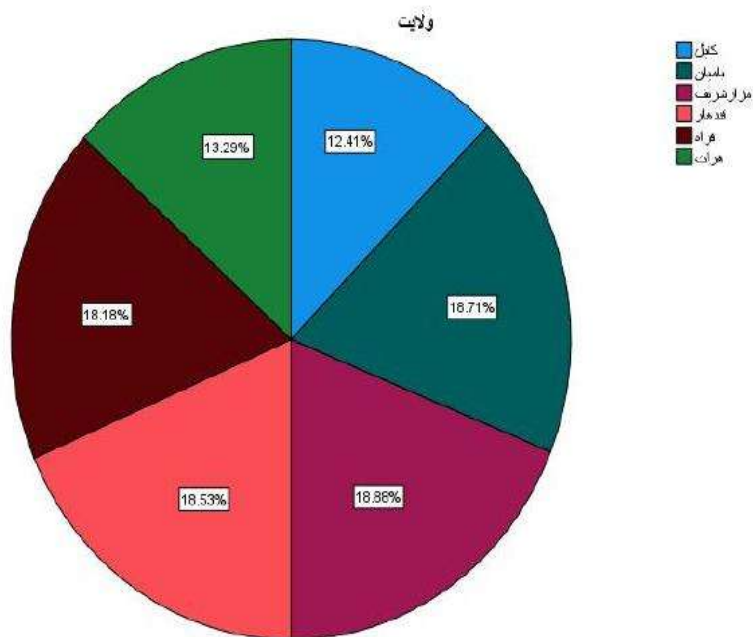


نمودار ۲. فراوانی سن نمونه آماری (معلمان و مدیران)



نمودار ۳. فراوانی سمت شغلی نمونه آماری

توزیع جغرافیایی نمونه نشان داد شرکت کنندگان از ولایت‌های کابل (۱۲.۴٪، $n = ۷۱$)، بامیان (۱۸.۷٪، $n = ۱۰۷$)، مزارشریف (۱۰.۸٪، $n = ۱۰۸$)، قندهار (۱۸.۵٪، $n = ۱۰۶$)، فراه (۱۸.۲٪، $n = ۱۰۴$) و هرات (۱۳.۳٪، $n = ۷۶$) بودند (نمودار ۴)؛ این توزیع تقریباً متوازن، امکان مقایسه بین ولایتی داده‌ها را فراهم آورد.



نمودار ۴. فراوانی ولایت محل خدمت نمونه آماری

در مجموع، بررسی ویژگی‌های جمعیت‌شناختی نشان می‌دهد نمونه پژوهش از نظر جنسیت، سن و توزیع جغرافیایی از تنوع قابل‌قبولی برخوردار بوده و در نتیجه، یافته‌های حاصل از تحلیل داده‌ها را می‌توان با احتیاط لازم به جامعه معلمان، مدیران و کارشناسان آموزش ابتدایی شش ولایت مورد مطالعه تعمیم داد. غلبه نسبی گروه سنی جوان‌تر (۳۰ تا ۴۰ سال)، با بیش از نیمی از نمونه نیز می‌تواند بیانگر ترکیب سنی نسبتاً جوان نیروی انسانی شاغل در آموزش ابتدایی افغانستان باشد که در تفسیر یافته‌های مربوط به اثر سن باید مورد توجه قرار گیرد. شاخص‌های آمار توصیفی مربوط به مقیاس کل و خرده‌مقیاس‌های پرسش‌نامه کارشناسان در جدول ۱ ارائه شده است.

جدول ۱. شاخص‌های آمار توصیفی پرسش‌نامه کارشناسان (n=۱۵۳)

مؤلفه	میانگین	انحراف استاندارد	خطای استاندارد میانگین	کمینه	بیشینه
طراحی و تدوین	۳/۵۷	۰/۲۷	۰/۰۲	-	-
ساخت	۳/۵۲	۰/۳۰	۰/۰۲	۲/۸۰	۴/۳۰
پیشنهادهای راهکارها	۳/۳۵	۰/۳۸	۰/۰۳	۲/۶۰	۴/۳۰
مقیاس کل	۳/۸۵	۰/۵۱	۰/۰۴	۲/۰۰	۴/۹۰

همچنین، شاخص‌های آمار توصیفی پرسش‌نامه معلمان و مدیران در جدول ۲ ارائه شده است.

جدول ۲. شاخص‌های آمار توصیفی پرسش‌نامه معلمان و مدیران (n=۴۱۹)

مؤلفه	میانگین	انحراف استاندارد	خطای استاندارد میانگین	کمینه	بیشینه
طراحی و تدوین	۳/۵۷	۰/۴۰	۰/۰۲	۲/۰۰	۴/۶۰
اجرا	۳/۸۲	۰/۳۹	۰/۰۲	۲/۶۰	۵/۰۰
ساخت	۳/۵۶	۰/۴۵	۰/۰۲	۲/۴۰	۴/۸۰
ارزشیابی	۳/۵۷	۰/۴۶	۰/۰۲	۲/۱۰	۴/۷۰

پیشنهادهای و راهکارها	۴/۰۶	۰/۵۱	۰/۰۲	۲/۵۰	۵/۰۰
مقیاس کل	۳/۷۲	۰/۲۶	۰/۰۱	-	-

مقایسه میانگین‌های جداول ۱ و ۲ نشان می‌دهد مؤلفه پیشنهادها و راهکارها هم در پرسش‌نامه کارشناسان ($M = ۳.۳۵, SD = ۰.۳۸$) و هم در پرسش‌نامه معلمان ($M = ۴.۰۶, SD = ۰.۵۱$) بالاترین میانگین را داشته است. در مقابل، مؤلفه ساخت در هر دو پرسش‌نامه، یعنی در کارشناسان ($M = ۳.۵۲, SD = ۰.۳۰$) و معلمان ($M = ۳.۵۶, SD = ۰.۴۵$)، در زمره پایین‌ترین میانگین‌ها قرار گرفته است. این الگو نشان‌دهنده توافق نسبی هر دو گروه ذی‌نفع بر بحرانی‌تر بودن حوزه ساخت و تولید کتاب‌های درسی نسبت به سایر مؤلفه‌هاست.

مقایسه شاخص‌های پراکندگی در جداول ۱ و ۲ نیز نکته جالب‌توجهی را آشکار می‌سازد: خرده‌مقیاس پیشنهادها و راهکارها که بالاترین میانگین را در هر دو پرسش‌نامه دارد، هم‌زمان بالاترین انحراف استاندارد را نیز در هر دو گروه به خود اختصاص داده است. این الگو نشان می‌دهد اگرچه به‌طور میانگین موافقت بالایی با راهکارهای اصلاحی وجود دارد، پراکندگی دیدگاه‌ها درباره این راهکارها نیز بیشتر از سایر مؤلفه‌هاست؛ به بیان دیگر، در حالی که اکثریت مشارکت‌کنندگان با ضرورت اصلاح موافق‌اند، توافق کمتری درباره نوع دقیق یا اولویت‌بندی این راهکارها میان آنان وجود دارد. در مقابل، خرده‌مقیاس طراحی و تدوین با کمترین انحراف استاندارد در پرسش‌نامه کارشناسان ($M = ۳.۵۷, SD = ۰.۲۷$)، بیانگر اجماع نسبی بیشتر کارشناسان درباره وضعیت این مؤلفه است.

روایی محتوایی هر دو پرسش‌نامه با استفاده از نظرات ۹ نفر از متخصصان برنامه‌ریزی درسی بررسی شد. پس از اعمال اصلاحات لازم درباره وضوح عبارات و تناسب گویه‌ها با مؤلفه‌های نظری، نسبت روایی محتوایی ($CVR \geq ۰.۷۸$) و شاخص روایی محتوایی ($CVI > ۰.۷۹$) برای تمامی گویه‌ها در محدوده قابل قبول قرار گرفت. روایی همگرا نیز با شاخص میانگین واریانس استخراج‌شده (AVE) بررسی شد که نتایج آن در جدول ۳ ارائه شده است.

جدول ۳. روایی همگرا (AVE) و پایایی ترکیبی (CR) خرده‌مقیاس‌های دو پرسش‌نامه

پرسش‌نامه	خرده‌مقیاس	میانگین واریانس استخراج‌شده (AVE)	پایایی ترکیبی (CR)
معلمان	طراحی و تدوین	۰/۵۰	۰/۷۱
معلمان	اجرا	۰/۵۲	۰/۷۱
معلمان	ساخت	۰/۵۲	۰/۷۷
معلمان	ارزشیابی	۰/۵۵	۰/۷۳
معلمان	پیشنهادهای و راهکارها	۰/۵۴	۰/۷۲
معلمان	مقیاس کل	۰/۵۸	۰/۷۸
کارشناسان	طراحی و تدوین	۰/۵۹	۰/۷۵
کارشناسان	ساخت	۰/۵۳	۰/۷۷
کارشناسان	پیشنهادهای و راهکارها	۰/۵۳	۰/۷۸
کارشناسان	مقیاس کل	۰/۵۵	۰/۸۴

همان‌گونه که در جدول ۳ مشاهده می‌شود، میانگین واریانس استخراج‌شده تمامی خرده‌مقیاس‌های هر دو ابزار بالاتر از ملاک ۵۰/۰ است ($AVE > .50$) که روایی همگرایی آن‌ها را تأیید می‌کند (هیر، ۲۰۱۴). همچنین، مقادیر پایایی ترکیبی همگی بالاتر از ملاک ۷۰/۰ به دست آمد ($CR = .71-.84$) که بیانگر همسانی درونی قابل قبول ابزارها برای تحلیل‌های بعدی است.

به‌منظور بررسی روایی سازه هر دو پرسش‌نامه، تحلیل عاملی تأییدی مرتبه دوم با نرم‌افزار SmartPLS نسخه ۳ انجام شد. شاخص‌های برازش مدل پرسش‌نامه کارشناسان در محدوده قابل قبول قرار داشت ($NFI = .90$, $SRMR = .09$). بار عاملی خرده‌مقیاس‌های طراحی و تدوین (λ) ۰.۹۱ (=)، ساخت ($\lambda = .89$) و پیشنهادها ($\lambda = .64$) بر سازه برنامه درسی معنی‌دار بود ($ps < .001$). جزئیات برآورد بارهای عاملی گویه‌ها و سطوح معنی‌داری در جدول ۴ ارائه شده است.

جدول ۴. نتایج تحلیل عاملی مرتبه اول و دوم و پایایی گویه‌ها

عامل مرتبه دوم	عامل مرتبه اول	بار عاملی مرتبه اول	سطح معنی‌داری	گویه	بار عاملی گویه	سطح معنی‌داری	پایایی گویه
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۱	۰/۳۸	۰/۰۰	۰/۱۴
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۲	۰/۵۱	۰/۰۰	۰/۲۶
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۳	۰/۵۷	۰/۰۰	۰/۳۲
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۴	۰/۶۱	۰/۰۰	۰/۳۷
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۵	۰/۶۳	۰/۰۰	۰/۴۰
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۶	۰/۵۷	۰/۰۰	۰/۳۲
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۷	۰/۴۴	۰/۰۰	۰/۱۹
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۸	۰/۵۸	۰/۰۰	۰/۳۴
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۹	۰/۳۲	۰/۰۴	۰/۱۰
برنامه درسی	طراحی و تدوین	۰/۹۱	۰/۰۰	سؤال ۱۰	۰/۷۳	۰/۰۰	۰/۵۳
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۱	۰/۷۰	۰/۰۰	۰/۴۹
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۲	۰/۷۷	۰/۰۰	۰/۵۹
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۳	۰/۳۷	۰/۰۴	۰/۱۴
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۴	۰/۳۴	۰/۰۴	۰/۱۲
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۵	۰/۴۸	۰/۰۰	۰/۲۳
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۶	۰/۶۲	۰/۰۰	۰/۳۸
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۷	۰/۵۰	۰/۰۰	۰/۲۵
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۸	۰/۳۳	۰/۰۲	۰/۱۱
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۹	۰/۳۷	۰/۰۰	۰/۱۴
برنامه درسی	ساخت برنامه درسی	۰/۸۹	۰/۰۰	سؤال ۱۰	۰/۳۱	۰/۰۴	۰/۱۰
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۱	۰/۳۶	۰/۰۱	۰/۱۳
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۲	۰/۳۹	۰/۰۱	۰/۱۵
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۳	۰/۳۴	۰/۰۲	۰/۱۲
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۴	۰/۴۳	۰/۰۲	۰/۱۸
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۵	۰/۶۳	۰/۰۲	۰/۴۰
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۶	۰/۵۴	۰/۰۳	۰/۲۹
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۷	۰/۳۶	۰/۰۲	۰/۱۳
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۸	۰/۳۵	۰/۰۳	۰/۱۲
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۹	۰/۴۴	۰/۰۱	۰/۱۹
برنامه درسی	پیشنهادهات	۰/۶۴	۰/۰۰	سؤال ۱۰	۰/۴۱	۰/۰۱	۰/۱۷

همان گونه که در جدول ۴ مشاهده می شود، در سطح سازه مرتبه دوم، بار عاملی هر سه مؤلفه در سطح ۰۰۱/۰ معنی دار است. در سطح گویه ها نیز تمامی بارهای عاملی مثبت و معنی دار بودند ($ps \leq .04$, $\lambda = .31-.77$). بالاترین بار گویه ای مربوط به سؤال ۲ مؤلفه ساخت ($\lambda = .77$) و پایین ترین بار مربوط به سؤال ۱۰ همین مؤلفه ($\lambda = .31$) است. ستون پایایی گویه در جدول ۴ برابر با مجذور بار عاملی گویه است ($\lambda^2 = .59-.10$)؛ بنابراین سهم واریانس هر گویه که توسط سازه مربوط تبیین می شود، در همین دامنه قرار دارد. در مجموع، این الگو همراه با شاخص های برازش مدل ($SRMR = .09$, $NFI = .90$)، از ساختار عاملی پرسش نامه کارشناسان پشتیبانی می کند.

برای پرسش نامه معلمان نیز شاخص های برازش بیانگر برازش مطلوب مدل با داده ها بود ($SRMR = .08$, $NFI = .91$). بارهای عاملی خرده مقیاس های طراحی و تدوین ($\lambda = .74$)، اجرا ($\lambda = .58$)، ساخت ($\lambda = .79$)، ارزشیابی ($\lambda = .71$) و پیشنهادها ($\lambda = .32$) بر سازه کلی برنامه درسی همگی معنی دار بود ($ps < .001$). جزئیات برآورد بارهای عاملی گویه ها و سطوح معنی داری در جدول ۵ ارائه شده است. با توجه به معنی داری بارهای عاملی سطح اول و دوم و برازش قابل قبول شاخص های مدل، ساختار عاملی هر دو ابزار پشتیبانی شد.

جدول ۵. برآورد پارامترهای بارهای عاملی مدل تحلیل عاملی تأییدی مرتبه دوم پرسش نامه معلمان

عامل مرتبه دوم	عامل مرتبه اول	بار عاملی مرتبه اول	سطح معنی داری	گویه	بار عاملی گویه	سطح معنی داری	پایایی گویه
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۱	۰/۶۱	۰/۰۰	۰/۳۷
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۲	۰/۷۶	۰/۰۰	۰/۵۸
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۳	۰/۵۰	۰/۰۰	۰/۲۵
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۴	۰/۵۰	۰/۰۰	۰/۲۵
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۵	۰/۳۱	۰/۰۳	۰/۱۰
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۶	۰/۳۹	۰/۰۴	۰/۱۵
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۷	۰/۳۷	۰/۰۳	۰/۱۴
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۸	۰/۴۶	۰/۰۰	۰/۲۱
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۹	۰/۵۷	۰/۰۰	۰/۳۲
برنامه درسی	طراحی و تدوین	۰/۷۴	۰/۰۰	سؤال ۱۰	۰/۳۵	۰/۰۴	۰/۱۲
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۱	۰/۳۶	۰/۰۲	۰/۱۳
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۲	۰/۶۰	۰/۰۰	۰/۳۶
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۳	۰/۳۷	۰/۰۴	۰/۱۴
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۴	۰/۳۳	۰/۰۴	۰/۱۱
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۵	۰/۳۰	۰/۰۳	۰/۰۹
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۶	۰/۴۰	۰/۰۱	۰/۱۶
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۷	۰/۳۵	۰/۰۴	۰/۱۲
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۸	۰/۳۰	۰/۰۲	۰/۰۹
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۹	۰/۵۴	۰/۰۰	۰/۲۹
برنامه درسی	اجرا	۰/۵۸	۰/۰۰	سؤال ۱۰	۰/۳۱	۰/۰۴	۰/۱۰
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۱	۰/۵۳	۰/۰۰	۰/۲۸
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۲	۰/۵۸	۰/۰۰	۰/۳۴
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۳	۰/۵۲	۰/۰۰	۰/۲۷
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۴	۰/۳۸	۰/۰۰	۰/۱۴
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۵	۰/۴۳	۰/۰۰	۰/۱۸

برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۶	۰/۵۵	۰/۰۰	۰/۳۰
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۷	۰/۴۶	۰/۰۰	۰/۲۱
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۸	۰/۵۰	۰/۰۰	۰/۲۵
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۹	۰/۳۴	۰/۰۱	۰/۱۲
برنامه درسی	ساخت	۰/۷۹	۰/۰۰	سؤال ۱۰	۰/۴۶	۰/۰۰	۰/۲۱
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۱	۰/۳۸	۰/۰۴	۰/۱۴
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۲	۰/۴۸	۰/۰۰	۰/۲۳
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۳	۰/۵۱	۰/۰۰	۰/۲۶
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۴	۰/۲۷	۰/۰۰	۰/۰۹
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۵	۰/۳۶	۰/۰۰	۰/۱۳
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۶	۰/۴۴	۰/۰۰	۰/۱۹
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۷	۰/۵۸	۰/۰۰	۰/۳۴
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۸	۰/۶۳	۰/۰۰	۰/۴۰
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۹	۰/۶۳	۰/۰۰	۰/۴۰
برنامه درسی	ارزشیابی	۰/۷۱	۰/۰۰	سؤال ۱۰	۰/۴۸	۰/۰۰	۰/۲۳
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۱	۰/۳۰	۰/۰۳	۰/۰۹
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۲	۰/۳۵	۰/۰۴	۰/۱۲
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۳	۰/۳۰	۰/۰۳	۰/۰۹
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۴	۰/۷۰	۰/۰۳	۰/۴۹
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۵	۰/۳۹	۰/۰۴	۰/۱۵
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۶	۰/۶۷	۰/۰۴	۰/۴۵
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۷	۰/۳۲	۰/۰۴	۰/۱۰
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۸	۰/۴۷	۰/۰۲	۰/۲۲
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۹	۰/۳۹	۰/۰۲	۰/۱۵
برنامه درسی	پیشنهادهات	۰/۳۲	۰/۰۰	سؤال ۱۰	۰/۳۳	۰/۰۴	۰/۱۱

در جدول ۵، بار عاملی مرتبه دوم هر پنج خرده‌مقیاس در سطح ۰/۱۰۰ معنی‌دار است. در سطح گویه‌ها، تمامی بارهای عاملی مثبت و در سطح ۰/۵۰ معنی‌دار بودند ($ps \leq .04$, $\lambda = .27-.76$). بالاترین بار گویه‌ای مربوط به سؤال ۲ مؤلفه طراحی و تدوین ($\lambda = .76$) و پایین‌ترین بار مربوط به سؤال ۴ مؤلفه ارزشیابی ($\lambda = .27$) است. مقادیر ستون پایایی گویه که برابر با مجذور بارهای عاملی گویه‌هاست، در دامنه ۰/۹ تا ۵۸/۰ قرار دارد ($\lambda^2 = .09-.58$). از منظر سازه مرتبه دوم، ساخت بیشترین بار عاملی ($\lambda = .79$) و پیشنهادها کمترین بار عاملی ($\lambda = .32$) را داشت؛ بنابراین سهم مؤلفه‌های سازنده در بازنمایی سازه کلی برنامه درسی یکسان نیست. این نتایج، در کنار شاخص‌های برازش مدل (SRMR = .۹۱, NFI = .۰۸)، از برازش قابل قبول مدل پرسش‌نامه معلمان حکایت دارد.

شایان ذکر است که در این پژوهش، از روش حداقل مربعات جزئی (PLS-SEM) به جای مدل‌یابی معادلات ساختاری مبتنی بر کوواریانس (CB-SEM) استفاده شد؛ زیرا PLS-SEM برای نمونه‌های با توزیع غیرکاملاً نرمال و اهداف پیش‌بینی‌محور (نظیر برآورد بارهای عاملی مؤلفه‌های سازه‌ای پیچیده مانند برنامه درسی) مناسب‌تر است و پیش‌فرض‌های آماری کمتری در مقایسه با روش‌های مبتنی بر کوواریانس دارد (هیر، ۲۰۱۴). این انتخاب روش‌شناختی، امکان بررسی هم‌زمان روایی سازه و برآورد بارهای عاملی مرتبه دوم را برای هر دو ابزار محقق ساخته

این پژوهش فراهم آورد. در این چارچوب، گزارش همزمان بارهای عاملی گویه‌ها و بارهای عاملی مرتبه اول بر سازه مرتبه دوم، ارزیابی تفصیلی تر روایی سازه را فراهم می‌کند.

برای بررسی میانگین عملکرد کلی و خرده‌مقیاس‌های پرسش‌نامه کارشناسان در مقایسه با میانگین نظری ۳، آزمون t تک‌نمونه‌ای اجرا شد. نتایج در جدول ۶ ارائه شده است. یافته‌ها نشان داد میانگین مؤلفه طراحی و تدوین به‌طور معناداری بالاتر از میانگین نظری بود ($t(152) = 21.52, p < .01, MD = 0.52, 95\% CI [0.47, 0.57]$). همچنین میانگین مؤلفه ساخت ($t(152) = 11.23, p < .01, MD = 0.35, 95\% CI [0.28, 0.41]$)، مؤلفه پیشنهادها و راهکارها ($t(152) = 20.66, p < .01, MD = 0.85, 95\% CI [0.77, 0.93]$)، و مقیاس کل ($t(152) = 25.82, p < .01, MD = 0.57, 95\% CI [0.53, 0.61]$) به‌طور معناداری بالاتر از میانگین نظری بودند. در همه موارد، کران پایین و بالای فاصله اطمینان ۹۵٪ تفاوت میانگین مثبت بود؛ بنابراین از دیدگاه کارشناسان، میانگین عملکرد کلی برنامه درسی و هر سه مؤلفه آن با اطمینان بیش از ۹۹٪ به‌طور معناداری بالاتر از میانگین نظری است.

جدول ۶. نتایج آزمون t تک‌نمونه‌ای پرسش‌نامه کارشناسان (میانگین نظری = ۳)

مؤلفه	آماره t	درجه آزادی	سطح معنی‌داری	تفاوت میانگین	کران پایین	کران بالا
طراحی و تدوین	۲۱/۵۲	۱۵۲	$<0/01$	۰/۵۲	۰/۴۷	۰/۵۷
ساخت	۱۱/۲۳	۱۵۲	$<0/01$	۰/۳۵	۰/۲۸	۰/۴۱
پیشنهادها و راهکارها	۲۰/۶۶	۱۵۲	$<0/01$	۰/۸۵	۰/۷۷	۰/۹۳
مقیاس کل	۲۵/۸۲	۱۵۲	$<0/01$	۰/۵۷	۰/۵۳	۰/۶۱

برای مقایسه رتبه سه مؤلفه از دیدگاه کارشناسان، آزمون فریدمن اجرا شد. نتایج نشان داد تفاوت معناداری میان رتبه‌های مؤلفه‌ها وجود دارد ($\chi^2(2, N = 153) = 114.18, p < .001$). میانگین رتبه‌ها در جدول ۷ ارائه شده است. مؤلفه پیشنهادها و راهکارها بالاترین رتبه ($M = 1.96$) و مؤلفه ساخت پایین‌ترین رتبه ($M = 1.45$) را داشت، درحالی‌که رتبه طراحی و تدوین ($M = 1.96$) بین آن دو قرار گرفت و تفاوت آن با ساخت محسوس نبود.

جدول ۷. میانگین رتبه مؤلفه‌های پرسش‌نامه کارشناسان (آزمون فریدمن)

مؤلفه	میانگین رتبه
طراحی و تدوین	۱/۹۶
ساخت	۱/۴۵
پیشنهادها و راهکارها	۲/۵۹

نتایج جدول ۷ نشان می‌دهد مؤلفه پیشنهادها و راهکارها به‌روشنی بالاترین رتبه و مؤلفه ساخت پایین‌ترین رتبه را در میان سه مؤلفه به خود اختصاص داده است، درحالی‌که تفاوت میان مؤلفه‌های طراحی و تدوین و ساخت محسوس نیست.

برای بررسی میانگین عملکرد کلی و پنج خرده‌مقیاس پرسش‌نامه معلمان و مدیران در مقایسه با میانگین نظری ۳، آزمون t تک‌نمونه‌ای اجرا شد. نتایج در جدول ۸ ارائه شده است. یافته‌ها نشان داد از دیدگاه معلمان و مدیران، میانگین عملکرد کلی و هر پنج مؤلفه برنامه به‌طور معناداری بالاتر از میانگین نظری بود: طراحی و تدوین ($t(418) = 29.37, p < .01, MD = 0.57, 95\% CI [0.54, 0.61]$)، اجرا ($t(418) = 42.74, p < .01, MD = 0.82, 95\% CI [0.78, 0.85]$)، ساخت ($t(418) = 25.62, p < .01, MD = 0.56, 95\% CI [0.52, 0.61]$)، ارزشیابی ($t(418) = 25.43, p < .01, MD = 0.57, 95\% CI [0.53, 0.62]$)، پیشنهادها و راهکارها ($t(418) = 42.56, p < .01, MD = 1.06, 95\% CI [1.01, 1.10]$)، و مقیاس کل ($t(418) = 56.20, p < .01, MD = 0.72, 95\% CI [0.69, 0.74]$). بالاترین تفاوت میانگین مربوط به مؤلفه پیشنهادها و راهکارها ($MD = 1.06$) و پایین‌ترین مربوط به مؤلفه طراحی و تدوین ($MD = 0.57$) بود.

جدول ۸. نتایج آزمون t تک‌نمونه‌ای پرسش‌نامه معلمان و مدیران (میانگین نظری = ۳)

مؤلفه	آماره t	درجه آزادی	سطح معنی‌داری	تفاوت میانگین	کران پایین	کران بالا
طراحی و تدوین	۲۹/۳۷	۴۱۸	$<0/01$	۰/۵۷	۰/۵۴	۰/۶۱
اجرا	۴۲/۷۴	۴۱۸	$<0/01$	۰/۸۲	۰/۷۸	۰/۸۵
ساخت	۲۵/۶۲	۴۱۸	$<0/01$	۰/۵۶	۰/۵۲	۰/۶۱
ارزشیابی	۲۵/۴۳	۴۱۸	$<0/01$	۰/۵۷	۰/۵۳	۰/۶۲
پیشنهادها و راهکارها	۴۲/۵۶	۴۱۸	$<0/01$	۱/۰۶	۱/۰۱	۱/۱۰
مقیاس کل	۵۶/۲۰	۴۱۸	$<0/01$	۰/۷۲	۰/۶۹	۰/۷۴

نتیجه آزمون فریدمن برای پنج مؤلفه پرسش‌نامه معلمان و مدیران نشان داد تفاوت معناداری میان رتبه‌های مؤلفه‌ها وجود دارد ($\chi^2(4, N = 311.81, p < .001)$). میانگین رتبه‌ها در جدول ۹ ارائه شده است. مؤلفه پیشنهادها و راهکارها بالاترین رتبه ($M = 4.01$) و پس از آن مؤلفه اجرا ($M = 3.38$) قرار داشت؛ سه مؤلفه طراحی و تدوین ($M = 2.54$)، ساخت ($M = 2.56$) و ارزشیابی ($M = 2.52$) تفاوت چندانی با یکدیگر نداشتند و در پایین‌ترین سطح رتبه‌بندی جای گرفتند.

جدول ۹. میانگین رتبه مؤلفه‌های پرسش‌نامه معلمان و مدیران (آزمون فریدمن)

مؤلفه	میانگین رتبه
طراحی و تدوین	۲/۵۴
اجرا	۳/۳۸
ساخت	۲/۵۶
ارزشیابی	۲/۵۲
پیشنهادها و راهکارها	۴/۰۱

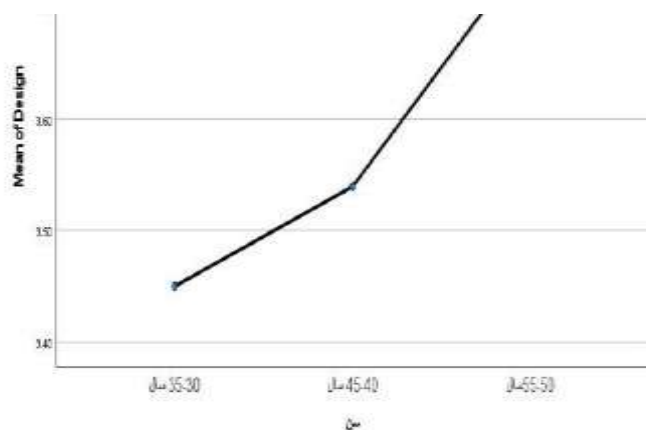
برای مقایسه دیدگاه دو گروه معلمان و کارشناسان درباره سه مؤلفه مشترک (طراحی و تدوین، ساخت، پیشنهادها)، آزمون t برای گروه‌های مستقل اجرا شد. نتایج آزمون لوین نشان داد مفروضه همسانی واریانس‌ها تنها برای مؤلفه پیشنهادها برقرار است (لوین برای طراحی: $F = 9.34$ ؛ ساخت: $F = 10.41$ ؛ پیشنهادها: $F = 0.61$). نتایج آزمون t در جدول ۱۰ ارائه شده است. در مؤلفه طراحی و تدوین تفاوت معناداری

میان دو گروه مشاهده نشد ($t(570) = -0.95, p = .34$). در مؤلفه ساخت تفاوت معنادار بود ($t(570) = -3.54, p < .01$) و در مؤلفه پیشنهادها نیز تفاوت معنادار بود ($t(570) = -2.95, p < .01$). بررسی میانگین‌های دو گروه نشان داد کارشناسان وضعیت ساخت و تولید کتاب‌های درسی را نامطلوب‌تر از معلمان ارزیابی کرده‌اند، درحالی‌که معلمان بیش از کارشناسان با راهکارهای اصلاحی پیشنهادی موافق بوده‌اند. این الگو را می‌توان ناشی از تفاوت زاویه دید دو گروه دانست: کارشناسان از منظر تخصصی‌تر به استانداردهای فنی تولید کتاب‌های درسی می‌نگرند، درحالی‌که معلمان به‌عنوان مجریان مستقیم برنامه درسی، نیاز فوری‌تری به اصلاحات عملی احساس می‌کنند.

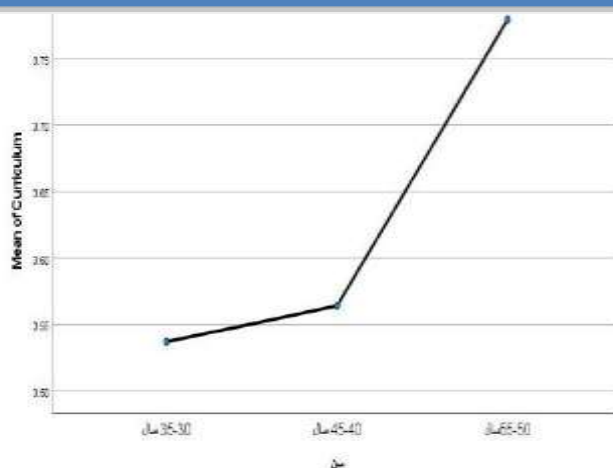
جدول ۱۰. نتایج آزمون t مستقل برای مقایسه دیدگاه معلمان و کارشناسان

مؤلفه	آماره F لوین	آماره t	درجه آزادی	سطح معنی‌داری
طراحی و تدوین	۹/۳۴	-۰/۹۵	۵۷۰	۰/۳۴
ساخت	۱۰/۴۱	-۳/۵۴	۵۷۰	<۰/۰۱
پیشنهادها و راهکارها	۰/۶۱	-۲/۹۵	۵۷۰	<۰/۰۱

همچنین همانطور که به‌طور خلاصه در نمودارهای شماره ۵ و ۶ ملاحظه می‌شوند، نتایج تحلیل واریانس یک‌راهه نشان داد سن شرکت‌کنندگان بر ارزیابی مؤلفه ساخت و مقیاس کل اثر معنادار دارد ($p < .05, \eta^2 = .05$)، هرچند اندازه اثر کوچک بود. آزمون تعقیبی شفه نشان داد این تفاوت عمدتاً میان گروه سنی ۵۰ تا ۵۵ سال با گروه‌های ۳۰ تا ۳۵ و ۴۰ تا ۴۵ سال معنادار است ($p < .05$)؛ به‌طوری‌که شرکت‌کنندگان مسن‌تر (۵۰ تا ۵۵ سال) ارزیابی مطلوب‌تری از مؤلفه ساخت و مقیاس کل ارائه کرده‌اند. روند میانگین این دو متغیر در گروه‌های سنی مختلف در نمودارهای ۷ و ۸ ارائه شده است.



نمودار ۵. میانگین خرده‌مقیاس ساخت به تفکیک گروه‌های سنی مختلف



نمودار ۶. میانگین مقیاس کل به تفکیک گروه‌های سنی مختلف

نتایج تحلیل واریانس یک‌راهه نشان داد سن شرکت‌کنندگان بر ارزیابی مؤلفه ساخت و مقیاس کل اثر معنادار داشت ($p < .05$, $\eta^2 = .05$). آزمون تعقیبی شفه نشان داد این تفاوت عمدتاً میان گروه سنی ۵۰ تا ۵۵ سال با گروه‌های ۳۰ تا ۳۵ و ۴۰ تا ۴۵ سال معنادار بود ($p < .05$); به‌طوری‌که شرکت‌کنندگان مسن‌تر (۵۰ تا ۵۵ سال) ارزیابی مطلوب‌تری از مؤلفه ساخت و مقیاس کل ارائه کردند. همچنین، سابقه خدمت و ولایت محل خدمت بر ارزیابی برخی مؤلفه‌ها، به‌ویژه مؤلفه ارزشیابی، اثر معنادار داشت ($p < .05$). تحلیل تکمیلی تفاوت‌های میان‌ولایتی نشان داد میانگین مؤلفه پیشنهادها و راهکارها در ولایت‌های هرات و مزارشریف به‌طور معناداری بالاتر از فراه و قندهار بود ($p < .05$). در مقابل، جنسیت در هیچ‌یک از مؤلفه‌ها اثر معنادار نشان نداد ($p > .05$) و سطح تحصیلات نیز در هیچ‌یک از مؤلفه‌های پنج‌گانه اثر معنادار نداشت ($p > .05$). به‌طور کلی، محور بررسی ۸ به‌طور جزئی تأیید شد؛ یعنی برخی متغیرهای جمعیت‌شناختی (سن، سابقه خدمت، ولایت) و نه همه آن‌ها (جنسیت، سطح تحصیلات)، بر ارزیابی مؤلفه‌های برنامه درسی اثرگذارند.

در مقابل، سطح تحصیلات شرکت‌کنندگان (دیپلم و کاردانی، کارشناسی، کارشناسی ارشد، دکتری) در هیچ‌یک از مؤلفه‌های پنج‌گانه اثر معنادار نداشت ($p > .05$). در تمامی موارد این یافته نشان می‌دهد ادراک معلمان و مدیران از عملکرد برنامه درسی، مستقل از سطح تحصیلات رسمی آنان است و بیشتر متأثر از تجربه عملی و میدانی (که با سن و سابقه خدمت مرتبط است) شکل می‌گیرد تا از دانش نظری اکتسابی در نظام آموزش عالی. این یافته پیامد عملی مهمی دارد: برنامه‌های توانمندسازی معلمان که صرفاً بر افزایش سطح تحصیلات رسمی (مانند ارتقای مدرک از دیپلم به کارشناسی) متمرکزند، ممکن است به‌تنهایی برای تغییر ادراک معلمان از کیفیت برنامه درسی کافی نباشند؛ در عوض، مداخلاتی که مستقیماً تجربه عملی و مهارت‌های کاربردی معلمان را هدف می‌گیرند (مانند کارگاه‌های روش تدریس و منتورینگ حین خدمت) احتمالاً اثربخشی بیشتری خواهند داشت.

تحلیل تکمیلی تفاوت‌های میان‌ولایتی نشان داد میانگین مؤلفه پیشنهادها و راهکارها در ولایت‌های هرات و مزارشریف به‌طور معناداری بالاتر از فراه و قندهار بود ($p < .05$) که می‌تواند بازتاب‌دهنده تفاوت در میزان دسترسی به منابع آموزشی تکمیلی و برنامه‌های حمایتی سازمان‌های غیردولتی فعال در مناطق شهری نسبت به مناطق کمتر برخوردار باشد. این یافته اهمیت لحاظ کردن بافت جغرافیایی و منطقه‌ای را در

سیاست‌گذاری‌های اصلاحی برنامه‌دستی برجسته می‌سازد؛ به این معنا که یک برنامه‌اصلاحی یکسان و متمرکز، ممکن است در همه ولایات به یک اندازه اثربخش نباشد. شایان ذکر است که ولایت‌های هرات و مزارشریف از نظر تاریخی به دلیل مجاورت با مرزهای تجاری و دسترسی نسبتاً بهتر به زیرساخت‌های ارتباطی، معمولاً از سطح بالاتری از فعالیت‌های سازمان‌های بین‌المللی و غیردولتی در بخش آموزش برخوردار بوده‌اند؛ این عامل زمینه‌ای می‌تواند بخشی از تفاوت مشاهده‌شده در ادراک شرکت‌کنندگان از راهکارهای اصلاحی و امکان اجرای آن‌ها را تبیین کند.

در نگاهی تلفیقی به یافته‌های پژوهش، می‌توان الگوی کلی زیر را استخراج کرد: (الف) عملکرد کلی برنامه‌دستی از دیدگاه هر دو گروه ذی‌نفع، مثبت و بالاتر از میانگین نظری ارزیابی شده است؛ (ب) در میان مؤلفه‌های چهارگانه، ساخت به‌طور مستمر پایین‌ترین جایگاه را دارد و پیشنهادها بالاترین جایگاه را کسب می‌کند؛ (ج) اختلاف‌نظر معنادار میان دو گروه ذی‌نفع محدود به مؤلفه‌هایی است که به تخصص فنی (ساخت) یا اولویت‌های عملیاتی (پیشنهادها) مربوط می‌شوند، نه به مؤلفه‌های راهبردی سطح بالا (طراحی)؛ و (د) عوامل زمینه‌ای مانند سن، سابقه خدمت و موقعیت جغرافیایی، نقش تعدیل‌کننده‌ای در نحوه ارزیابی ذی‌نفعان ایفا می‌کنند.

علاوه بر معناداری آماری، اندازه اثر d کوهن برای تفاوت‌های میانگین در آزمون t تک‌نمونه‌ای محاسبه شد. مقدار d برای مقیاس کل پرسش‌نامه معلمان برابر با ۰.۷۷ و برای مقیاس کل پرسش‌نامه کارشناسان برابر با ۱.۹۰ به دست آمد که هر دو مطابق معیار کوهن در دسته اندازه اثر بسیار بزرگ قرار می‌گیرند. این مقادیر بزرگ اندازه اثر، عمدتاً ناشی از حجم نمونه بزرگ پژوهش (۷۲۵ نفر) و انحراف استاندارد نسبتاً کوچک نمرات (به‌ویژه در مقیاس کل) است؛ بنابراین، معناداری آماری گسترده در تمامی آزمون‌های t تک‌نمونه‌ای را نباید به‌تنهایی نشانه قطعی مطلوبیت وضعیت موجود دانست.

در مقابل، اندازه اثر تفاوت میان دو گروه معلمان و کارشناسان در آزمون t مستقل (بر اساس d کوهن) برای مؤلفه ساخت برابر با ۰.۳۰ و برای مؤلفه پیشنهادها برابر با ۰.۲۵ بود که هر دو در دسته اندازه اثر کوچک تا متوسط قرار می‌گیرند. این تفاوت میان بزرگی اثر درون‌گروهی (نسبت به میانگین نظری) و بین‌گروهی (میان دو گروه ذی‌نفع) نکته روش‌شناختی مهمی را آشکار می‌سازد: اگرچه هر دو گروه به‌طور مطلق برنامه‌دستی را مطلوب‌تر از حد متوسط ارزیابی می‌کنند، تفاوت نسبی میان دیدگاه آنان - هرچند از نظر آماری معنادار - از نظر عملی چندان بزرگ نیست؛ بنابراین باید در تفسیر نتایج، به‌جای تمرکز صرف بر سطح معناداری (p)، به اندازه اثر نیز توجه شود تا از تفسیر اغراق‌آمیز یافته‌ها پرهیز شود.

بحث و نتیجه‌گیری

هدف پژوهش حاضر بررسی کمی عملکرد برنامه‌های درسی دوره ابتدایی در نظام آموزشی افغانستان از دیدگاه معلمان، مدیران، کارشناسان و متخصصان و مقایسه وضعیت مؤلفه‌های طراحی و تدوین، ساخت و تولید، اجرا، ارزشیابی و پیشنهادها و راهکارها بود. یافته‌های اصلی نشان داد که ارزیابی کلی هر دو گروه از عملکرد برنامه‌دستی به‌طور معنی‌داری بالاتر از میانگین نظری بود؛ به‌گونه‌ای که در گروه کارشناسان، عملکرد

کلی برنامه درسی با آماره $t=25.82$ و در گروه معلمان و مدیران با آماره $t=56.20$ به طور معنی داری بالاتر از سطح متوسط قرار گرفت. همچنین، تمامی مؤلفه‌های مورد بررسی در هر دو گروه از میانگین نظری ۳ بالاتر بودند. با این حال، نتایج آزمون فریدمن نشان داد که میان مؤلفه‌های مختلف تفاوت معنی داری وجود دارد و این بدان معناست که مطلوب ارزیابی شدن عملکرد کلی برنامه درسی نباید به معنای عملکرد یکسان و متوازن همه اجزای آن تلقی شود. در گروه معلمان و مدیران، پیشنهادها و راهکارها بالاترین رتبه را به دست آورد و پس از آن اجرا قرار گرفت، در حالی که طراحی و تدوین، ساخت و ارزشیابی در رتبه‌های پایین تر قرار داشتند. در گروه کارشناسان نیز پیشنهادها و راهکارها در جایگاه بالاتر و ساخت در جایگاه پایین تر قرار گرفت. افزون بر این، مقایسه دو گروه نشان داد در مؤلفه طراحی و تدوین تفاوت معنی داری میان دیدگاه آنان وجود نداشت، اما در مؤلفه‌های ساخت و پیشنهادها تفاوت معنی دار بود و کارشناسان وضعیت ساخت و تولید کتاب‌های درسی را انتقادی تر از معلمان ارزیابی کردند. همچنین، سن، سابقه خدمت و ولایت محل خدمت با ارزیابی برخی مؤلفه‌ها ارتباط معنی دار داشتند، در حالی که جنسیت و سطح تحصیلات اثر معنی داری نشان ندادند.

بالاتر بودن عملکرد کلی برنامه درسی از سطح متوسط را می‌توان نشانه وجود ظرفیت‌های قابل قبول در ساختار رسمی برنامه درسی دوره ابتدایی افغانستان دانست. این یافته بیانگر آن است که مشارکت کنندگان، علی‌رغم وجود محدودیت‌های شناخته شده در نظام آموزشی افغانستان، کل برنامه درسی را نامطلوب ارزیابی نکرده‌اند و در ابعادی مانند تعیین اهداف، سازمان‌دهی محتوا، اجرای آموزشی و برخی سازوکارهای ارزشیابی، عناصر مثبتی را تجربه کرده‌اند. این نتیجه با دیدگاه‌هایی همسو است که برنامه درسی را مجموعه‌ای چندبعدی و پویا می‌دانند که کیفیت آن حاصل تعامل میان طراحی، محتوا، روش اجرا و ارزشیابی است (Fathi Vajargah, 2023; Maleki, 2019). از منظر نظری، مطلوب بودن نسبی عملکرد کلی می‌تواند ناشی از آن باشد که طی سال‌های گذشته، تلاش‌هایی برای سازمان‌دهی رسمی تر اهداف، محتوای آموزشی و ساختار کتاب‌های درسی در افغانستان صورت گرفته است. با این حال، همان گونه که فولان تأکید می‌کند، تغییر آموزشی پایدار زمانی تحقق می‌یابد که اجزای مختلف نظام آموزشی به صورت هم‌راستا عمل کنند و وجود پیشرفت در یک یا چند بخش نمی‌تواند به تنهایی تضمین کننده اثربخشی کل نظام باشد (Fullan, 2016). بنابراین، بالاتر بودن میانگین کلی از نقطه متوسط را باید به عنوان نشانه وجود ظرفیت نسبی تلقی کرد، نه دلیلی بر نبود مسئله یا نیاز به اصلاح.

در عین حال، تفاوت معنی دار میان مؤلفه‌های برنامه درسی یکی از مهم ترین یافته‌های پژوهش است و نشان می‌دهد که کیفیت عملکرد در مراحل مختلف چرخه برنامه درسی یکنواخت نیست. این یافته با چارچوب‌های نظری برنامه ریزی درسی هماهنگ است که بر لزوم انسجام میان اهداف، محتوا، مواد آموزشی، اجرا و ارزشیابی تأکید دارند (Marsh & Willis, 2007; Ornstein & Hunkins, 2018). اگرچه طراحی رسمی یک برنامه ممکن است نسبتاً مناسب باشد، کیفیت نهایی آن به چگونگی تبدیل این طراحی به مواد آموزشی، امکانات مدرسه‌ای، روش‌های تدریس و سازوکارهای ارزشیابی وابسته است. از این رو، فاصله میان رتبه مؤلفه‌ها می‌تواند نشانه‌ای از شکاف میان «برنامه قصدشده» و «برنامه اجراشده» باشد. این تفسیر با یافته‌های Phaeton و Stears همخوان است که نشان دادند میان برنامه رسمی و نحوه تفسیر و اجرای آن توسط معلمان ممکن است تفاوت معناداری شکل گیرد و درک معلم از اهداف و الزامات برنامه نقشی تعیین کننده در تبدیل سیاست

درسی به عمل آموزشی دارد (Phaeton & Stears, 2016). در نظام‌های آموزشی دارای محدودیت منابع، این شکاف ممکن است شدیدتر باشد؛ زیرا حتی محتوای نسبتاً مناسب نیز در صورت فقدان زیرساخت، آموزش حرفه‌ای و منابع کافی به صورت کامل اجرا نمی‌شود. پایین‌تر بودن جایگاه مؤلفه ساخت و تولید، به‌ویژه از دیدگاه کارشناسان، اهمیت زیادی دارد. مواد آموزشی و کتاب‌های درسی یکی از اصلی‌ترین واسطه‌های تبدیل اهداف برنامه درسی به تجربه‌های واقعی یادگیری هستند و کیفیت فنی و محتوایی آن‌ها می‌تواند مستقیماً بر فرصت یادگیری دانش‌آموزان اثر بگذارد. یافته حاضر با نتایج Sievert و همکاران هماهنگ است که نشان دادند کیفیت کتاب‌های درسی می‌تواند با عملکرد دانش‌آموزان در تکالیف آموزشی ارتباط داشته باشد و چگونگی سازمان‌دهی مطالب، وضوح ارائه و نوع فعالیت‌های پیشنهادی در شکل‌دهی به فرصت‌های یادگیری نقش دارد (Sievert et al., 2021). در افغانستان، اهمیت کتاب درسی احتمالاً بیشتر نیز هست؛ زیرا در بسیاری از مدارس، به‌ویژه در مناطق برخوردار از منابع محدود، کتاب درسی اصلی‌ترین منبع سازمان‌یافته یاددهی و یادگیری محسوب می‌شود. بنابراین، کاستی در کیفیت چاپ، صحافی، تصاویر، تناسب محتوای بصری و نوشتاری یا در دسترسی بودن به موقع کتاب‌ها می‌تواند به یکی از گلوگاه‌های اصلی اجرای برنامه درسی تبدیل شود. گزارش‌های بین‌المللی نیز به محدودیت منابع آموزشی و دشواری‌های تأمین فرصت‌های یادگیری باکیفیت در افغانستان اشاره کرده‌اند (Unesco, 2023; Unicef, 2025). از این منظر، یافته پژوهش حاضر نشان می‌دهد که ارتقای کیفیت برنامه درسی بدون اصلاح فرآیند تولید، کنترل کیفیت و توزیع مواد آموزشی دشوار خواهد بود.

اینکه کارشناسان وضعیت مؤلفه ساخت را نامطلوب‌تر از معلمان ارزیابی کرده‌اند، می‌تواند ناشی از تفاوت جایگاه حرفه‌ای و معیارهای قضاوت دو گروه باشد. کارشناسان و متخصصان آموزشی معمولاً با استانداردهای فنی تولید کتاب، اصول طراحی محتوا، هماهنگی عمودی و افقی برنامه و معیارهای رسمی کیفیت آشنایی بیشتری دارند و بنابراین ممکن است کاستی‌هایی را تشخیص دهند که برای معلم در جریان فعالیت روزمره کلاس کمتر برجسته باشد. در مقابل، معلمان غالباً کتاب و مواد آموزشی را از منظر قابلیت استفاده عملی، انطباق با شرایط کلاس و امکان تدریس ارزیابی می‌کنند. این تفاوت زاویه دید با منطق ارزشیابی چندمنبعی همخوان است که بر گردآوری دیدگاه گروه‌های مختلف برای دستیابی به تصویری کامل‌تر از عملکرد برنامه تأکید می‌کند (Patton, 2008). نبود تفاوت معنی‌دار در مؤلفه طراحی و تدوین نیز می‌تواند نشان دهد که دو گروه درباره کیفیت نسبی چارچوب رسمی برنامه توافق بیشتری دارند، درحالی‌که در مرحله تبدیل طراحی به محصول آموزشی، اختلاف دیدگاه آشکارتر می‌شود. این الگو از این فرض پشتیبانی می‌کند که بسیاری از مسائل برنامه درسی نه الزاماً در سطح اهداف رسمی، بلکه در فرایند عملیاتی‌سازی، تولید و اجرای آن پدیدار می‌شوند.

کسب بالاترین رتبه توسط مؤلفه پیشنهادها و راهکارها نیز از نظر تفسیری مهم است. بالا بودن این مؤلفه الزاماً نشان‌دهنده رضایت بیشتر از وضعیت موجود نیست، بلکه می‌تواند بیانگر توافق نسبتاً گسترده مشارکت‌کنندگان با ضرورت اصلاح و بهبود برنامه باشد. درواقع، وقتی معلمان، مدیران و کارشناسان با پیشنهادهایی مانند بازنگری محتوا، بهبود منابع آموزشی، تقویت آموزش‌های حرفه‌ای و اصلاح سازوکارهای ارزشیابی موافقت بیشتری نشان می‌دهند، این امر می‌تواند نشان‌دهنده آگاهی مشترک آنان از وجود فاصله میان وضعیت موجود و وضعیت مطلوب باشد. این یافته با دیدگاه Fullan سازگار است که اصلاح آموزشی را فرایندی نیازمند مشارکت فعال ذی‌نفعان و ایجاد ظرفیت در سطح مدرسه و

نظام آموزشی می‌داند (Fullan, 2016). همچنین، رویکرد ارزشیابی استفاده‌محور تأکید می‌کند که ارزیابی زمانی ارزشمند است که نتایج آن بتواند مستقیماً در تصمیم‌گیری و اصلاح برنامه به کار گرفته شود (Patton, 2008). بنابراین، حمایت بالای ذی‌نفعان از راهکارهای اصلاحی می‌تواند فرصتی سیاستی تلقی شود؛ زیرا نشان می‌دهد زمینه ادراکی مناسبی برای پذیرش تغییرات وجود دارد.

رتبه نسبتاً بالاتر مؤلفه اجرا در میان معلمان و مدیران در مقایسه با طراحی، ساخت و ارزشیابی نیز قابل توجه است. این نتیجه ممکن است منعکس‌کننده توان سازگارشدن معلمان با محدودیت‌های موجود و استفاده از راهبردهای عملی برای اجرای برنامه باشد. معلمان در بسیاری از محیط‌های آموزشی مجبورند با کمبود منابع، تفاوت سطح دانش‌آموزان و شرایط متغیر مدرسه سازگار شوند و در نتیجه ممکن است عملکرد اجرایی خود را نسبتاً مطلوب ارزیابی کنند، حتی اگر مواد آموزشی یا امکانات کاملاً مطلوب نباشد. با این حال، مطالعات مربوط به اجرای برنامه درسی نشان داده‌اند که کیفیت اجرا به میزان زیادی به تفسیر معلم از برنامه و ظرفیت او برای انطباق اهداف رسمی با واقعیت کلاس وابسته است (Phaeton & Stears, 2016). پژوهش Parmigiani و همکاران نیز نشان می‌دهد که توسعه حرفه‌ای معلمان، به‌ویژه آموزش مبتنی بر پژوهش عملی و ارزشیابی تکوینی، می‌تواند کیفیت اجرای برنامه را بهبود بخشد (Parmigiani et al., 2025). بنابراین، مطلوب بودن نسبی اجرا نباید باعث نادیده گرفتن ضرورت سرمایه‌گذاری بر توانمندسازی مستمر معلمان شود، به‌ویژه در شرایطی که برنامه‌های درسی با نیازهای نوظهور و فناوری‌های جدید مواجه‌اند (Song, 2025).

جایگاه نسبتاً پایین‌تر ارزشیابی در میان معلمان و مدیران نیز نشان می‌دهد که سنجش یادگیری ممکن است یکی از نقاط نیازمند بهبود در چرخه برنامه درسی باشد. ارزشیابی در رویکردهای جدید صرفاً ابزاری برای نمره‌دهی و طبقه‌بندی دانش‌آموزان نیست، بلکه باید به‌عنوان منبع بازخورد برای بهبود تدریس و یادگیری عمل کند. Wiliam تأکید می‌کند که ارزشیابی برای یادگیری زمانی اثربخش است که اطلاعات حاصل از آن به صورت مستمر در تصمیم‌های آموزشی معلم وارد شود و به دانش‌آموز امکان دهد مسیر یادگیری خود را اصلاح کند (Wiliam, 2011). از این منظر، پایین‌تر بودن رتبه ارزشیابی می‌تواند ناشی از غلبه روش‌های سنتی‌تر، محدودیت در ابزارهای سنجش عملکردی، کمبود آموزش تخصصی یا دشواری اجرای ارزشیابی مستمر در کلاس‌های پرجمعیت باشد. یافته‌های Parmigiani و همکاران نیز نشان می‌دهد که آموزش هدفمند معلمان می‌تواند اجرای ارزشیابی تکوینی را تقویت کند (Parmigiani et al., 2025). از این رو، ارتقای برنامه درسی افغانستان نیازمند آن است که اصلاح ارزشیابی هم‌زمان با بازنگری محتوا و آموزش معلمان پیش رود.

یافته دیگر پژوهش، اثر معنی‌دار سن، سابقه خدمت و ولایت محل خدمت بر ارزیابی برخی مؤلفه‌های برنامه بود، درحالی‌که جنسیت و سطح تحصیلات اثر معنی‌داری نداشتند. تفاوت برحسب سن و سابقه می‌تواند نشان‌دهنده نقش تجربه حرفه‌ای و مقایسه تاریخی در ارزیابی برنامه باشد؛ افراد با سابقه‌تر ممکن است وضعیت موجود را در مقایسه با دوره‌های پیشین بسنجند، درحالی‌که نیروهای جوان‌تر بر اساس انتظارات جدیدتر از آموزش، فناوری و استانداردهای معاصر قضاوت کنند. ادبیات برنامه‌ریزی آموزشی نیز نشان می‌دهد که برداشت افراد از اثربخشی برنامه متأثر از تجربه کاری، زمینه سازمانی و میزان مواجهه آنان با تغییرات آموزشی است (Mohsenpour, 2021; Taghipour Zahir, 2022). تفاوت میان ولایت‌ها نیز می‌تواند بازتاب نابرابری در دسترسی به منابع، زیرساخت، نیروی انسانی، توزیع کتاب، فرصت‌های آموزش

حرفه‌ای و شرایط اجرایی باشد. گزارش‌های یونسکو و یونیسف بر وجود تفاوت‌های قابل توجه در دسترسی و کیفیت فرصت‌های آموزشی در بافت‌های مختلف افغانستان تأکید کرده‌اند (Unesco, 2023; Unicef, 2025). در مقابل، نبود اثر معنی‌دار جنسیت و تحصیلات می‌تواند نشان دهد که ارزیابی برنامه بیش از آنکه تابع این ویژگی‌های فردی باشد، تحت تأثیر تجربه حرفه‌ای و شرایط محل خدمت قرار دارد.

یافته‌های پژوهش حاضر در بستر گسترده‌تر آموزش افغانستان نیز معنادار است. مطالعات تاریخی نشان داده‌اند که نظام آموزشی افغانستان طی دهه‌های مختلف تحت تأثیر تحولات سیاسی، اجتماعی و نهادی قرار داشته و ثبات برنامه‌ریزی آموزشی همواره با چالش مواجه بوده است (Andishmand, 2010). تجربه نظام‌های آموزشی منطقه نیز نشان می‌دهد که تحولات ساختاری و سیاسی می‌توانند به‌طور مستقیم بر سازمان آموزش و محتوای درسی اثرگذار باشند (Amiri, 2020). Aghazadeh, 2020 نیز در تحلیل برنامه درسی ملی مدارس ابتدایی افغانستان پس از ۱۱ سپتامبر به شکاف‌هایی میان جهت‌گیری‌های رسمی و محتوای واقعی برنامه اشاره کرده است (Amiri, 2016). بنابراین، یافته حاضر مبنی بر وجود تفاوت معنی‌دار میان مؤلفه‌های برنامه را می‌توان در امتداد همین مسئله تاریخی تفسیر کرد: ساختار رسمی ممکن است در برخی ابعاد تقویت شده باشد، اما تبدیل آن به مواد آموزشی باکیفیت، اجرای همگون و ارزشیابی اثربخش همچنان با چالش همراه است.

از سوی دیگر، تحولات جهانی در آموزش نشان می‌دهد که معیارهای کیفیت برنامه درسی نیز در حال تغییر است. نظام‌های آموزشی امروز علاوه بر دانش پایه، با ضرورت توجه به مهارت‌های شناختی، اجتماعی، سلامت، توان سازگاری و شایستگی‌های دیجیتال مواجه‌اند. Benes و Alperin بر ضرورت هماهنگی میان محتوا، تدریس و ارزشیابی در طراحی برنامه‌های آموزشی اثربخش تأکید می‌کنند (Benes & Alperin, 2025). Mallillin و Mallillin توجه به ویژگی‌های روان‌شناختی و رفتاری دانش‌آموزان را در تدوین برنامه‌های توسعه مدرسه ضروری می‌دانند (Mallillin & Mallillin, 2024). همچنین، برنامه درسی پنهان و محیط اجتماعی مدرسه می‌تواند بر شکل‌گیری شخصیت و رفتار یادگیرندگان اثرگذار باشد (Victor Ojorka et al., 2024). بنابراین، اصلاح برنامه درسی ابتدایی افغانستان نباید صرفاً به تغییر متن کتاب محدود شود، بلکه باید کیفیت کل تجربه آموزشی کودک، از محیط مدرسه و روش تدریس تا ارزشیابی و مهارت‌های نوظهور را در بر گیرد. نتایج بین‌المللی مربوط به یادگیری پس از اختلال‌های گسترده آموزشی نیز نشان داده‌اند که ارتقای کیفیت و بازیابی یادگیری نیازمند رویکردی نظام‌مند و چندبعدی است (Oecd, 2023).

در مجموع، نتایج پژوهش نشان می‌دهد که برنامه درسی دوره ابتدایی افغانستان از نگاه ذی‌نفعان وضعیت کاملاً نامطلوبی ندارد و عملکرد کلی آن بالاتر از سطح متوسط ارزیابی می‌شود، اما این تصویر کلی ناهمگونی قابل توجه میان اجزای برنامه را پنهان می‌کند. بیشترین زمینه برای اصلاح در حوزه ساخت و تولید مواد آموزشی، ارزشیابی و برخی ابعاد اجرای برنامه مشاهده می‌شود و توافق بالای مشارکت‌کنندگان با پیشنهادها و اصلاحی نیز مؤید آمادگی آنان برای تغییر است. بر اساس اصول برنامه‌ریزی درسی، بهبود پایدار تنها زمانی حاصل می‌شود که اصلاحات در طراحی، تولید، اجرا و ارزشیابی به صورت یکپارچه انجام گیرد (Marsh & Willis, 2007; Ornstein & Hunkins, 2018). در سطح روش‌شناختی نیز استفاده از ابزارهای دارای ساختار اندازه‌گیری معتبر برای پایش مستمر برنامه اهمیت دارد و توجه به روایی محتوا و کیفیت مدل اندازه‌گیری می‌تواند اعتمادپذیری تصمیم‌های مبتنی بر داده را افزایش دهد (Hair, 2014; Polit & Beck, 2006).

از این منظر، پژوهش حاضر نشان می‌دهد که سیاست‌گذاری آینده در آموزش ابتدایی افغانستان باید از رویکرد اصلاحات پراکنده فاصله گرفته و بر یک زنجیره منسجم از طراحی تا اجرا و ارزشیابی تمرکز کند.

این پژوهش با چند محدودیت همراه بود که در تفسیر و تعمیم نتایج باید مورد توجه قرار گیرد. نخست، داده‌ها بر اساس خودگزارشی معلمان، مدیران، کارشناسان و متخصصان گردآوری شد و بنابراین احتمال سوگیری پاسخ‌دهی، تأثیر انتظارات سازمانی یا ارائه تصویری مطلوب‌تر از وضعیت واقعی وجود دارد. هرچند محرمانگی پاسخ‌ها تضمین شد، جایگاه اداری برخی شرکت‌کنندگان ممکن است همچنان بر نحوه ارزیابی آنان اثر گذاشته باشد. دوم، مطالعه به شش ولایت کابل، بامیان، مزارشریف، قندهار، فراه و هرات محدود بود؛ از این رو، تعمیم نتایج به تمام مناطق افغانستان، به‌ویژه مناطق دورافتاده با شرایط آموزشی متفاوت، باید با احتیاط انجام شود. سوم، طرح پژوهش مقطعی بود و از این رو تغییرات عملکرد برنامه درسی در طول زمان یا اثر اصلاحات بعدی قابل بررسی نبود. چهارم، دو پرسش‌نامه متفاوت برای معلمان و کارشناسان استفاده شد و در نتیجه مقایسه مستقیم دو گروه تنها در سه مؤلفه مشترک امکان‌پذیر بود. همچنین، سنجش عملکرد برنامه درسی عمدتاً بر ادراک ذی‌نفعان استوار بود و شاخص‌های عینی‌تر مانند مشاهده کلاس، تحلیل کیفیت واقعی کتاب‌های درسی، عملکرد تحصیلی دانش‌آموزان و میزان دسترسی مدارس به منابع آموزشی به‌طور مستقیم وارد تحلیل نشدند.

پیشنهاد می‌شود پژوهش‌های آینده با استفاده از طرح‌های ترکیبی، یافته‌های کمی حاضر را با مصاحبه‌های عمیق، گروه‌های کانونی، مشاهده کلاس و تحلیل محتوای کتاب‌های درسی تکمیل کنند تا دلایل تفاوت میان مؤلفه‌ها و گروه‌های ذی‌نفع با دقت بیشتری شناسایی شود. انجام مطالعات طولی نیز می‌تواند تغییرات عملکرد برنامه درسی را در پی اجرای اصلاحات آموزشی بررسی کند. گسترش نمونه به ولایت‌های بیشتر و به‌ویژه مناطق روستایی، مرزی و کم‌برخوردار می‌تواند تصویر جامع‌تری از نابرابری‌های جغرافیایی فراهم سازد. همچنین پیشنهاد می‌شود رابطه کیفیت ادراک‌شده مؤلفه‌های برنامه با پیامدهای واقعی دانش‌آموزان، از جمله سواد خواندن، ریاضیات، حضور در مدرسه، انگیزش و مشارکت تحصیلی، بررسی شود. مطالعات آینده می‌توانند از مدل‌سازی معادلات ساختاری برای آزمون روابط میان کیفیت طراحی، تولید، اجرا و ارزشیابی و نیز نقش میانجی یا تعدیل‌کننده متغیرهایی مانند آموزش حرفه‌ای معلمان، دسترسی به منابع و شرایط مدرسه استفاده کنند. مقایسه افغانستان با کشورهای دارای شرایط اقتصادی، فرهنگی یا پس از منازعه مشابه نیز می‌تواند به شناسایی راهکارهای قابل انتقال برای اصلاح برنامه درسی کمک کند.

در سطح اجرایی، پیشنهاد می‌شود وزارت معارف و نهادهای مسئول برنامه‌ریزی درسی بازنگری نظام‌مند فرآیند تولید کتاب‌های درسی را در اولویت قرار دهند و سازوکارهای مشخصی برای کنترل کیفیت محتوایی، زبانی، بصری و فنی پیش از چاپ و توزیع ایجاد کنند. مشارکت سازمان‌یافته معلمان، مدیران و کارشناسان استانی در بازنگری برنامه می‌تواند فاصله میان طراحی مرکزی و واقعیت مدرسه را کاهش دهد. همچنین لازم است برنامه‌های منظم توسعه حرفه‌ای برای معلمان با تمرکز بر اجرای انعطاف‌پذیر برنامه درسی، روش‌های فعال تدریس، ارزشیابی تکوینی و استفاده اثربخش از منابع محدود طراحی شود. ایجاد نظام بازخورد مستمر که در آن مدارس بتوانند مشکلات کتاب‌ها، دشواری‌های اجرا و نیازهای دانش‌آموزان را مستقیماً به سطوح تصمیم‌گیری منتقل کنند، می‌تواند اصلاحات را از حالت دوره‌ای و مقطعی به فرآیندی پیوسته

تبدیل کند. در کنار آن، تخصیص منابع باید با توجه به تفاوت‌های میان ولایت‌ها و میزان محرومیت آموزشی صورت گیرد تا کیفیت اجرای برنامه درسی کمتر به محل جغرافیایی مدرسه وابسته باشد. در نهایت، پیشنهاد می‌شود ارزیابی دوره‌ای عملکرد برنامه درسی با استفاده از شاخص‌های استاندارد و ترکیبی از داده‌های معلمان، کارشناسان، مشاهده کلاس و عملکرد دانش‌آموزان به بخشی ثابت از نظام تضمین کیفیت آموزش ابتدایی افغانستان تبدیل شود.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ‌گونه تضاد منافی وجود ندارد.

موازین اخلاقی

در تمامی مراحل پژوهش حاضر اصول اخلاقی مرتبط با نشر و انجام پژوهش رعایت گردیده است.

تشکر و قدردانی

از تمامی کسانی که در انجام این پژوهش ما را همراهی کردند تشکر و قدردانی به عمل می‌آید.

EXTENDED SUMMARY

Introduction

Primary education constitutes the foundational stage of formal schooling and plays a decisive role in shaping children's cognitive, social, emotional, and behavioral development. At this stage, students acquire the basic competencies of reading, writing, numeracy, communication, problem solving, and self-regulated learning that influence their subsequent educational trajectories. Within this process, curriculum is not merely a collection of textbooks or prescribed subject matter; rather, it is a systematic framework through which educational goals are defined, content is selected and organized, learning opportunities are designed, instructional activities are implemented, and learning outcomes are assessed. Accordingly, the quality of a curriculum depends on the coherence among its major components, including design, content production, implementation, and evaluation (Fathi Vajargah, 2023; Maleki, 2019). Contemporary curriculum theory similarly emphasizes that effective curriculum planning requires alignment among intended objectives, instructional materials, classroom implementation, and assessment processes (Marsh & Willis, 2007; Ornstein & Hunkins, 2018). This alignment has become increasingly important as educational systems are expected to respond not only to conventional literacy needs but also to emerging competencies associated with technology, digital learning, and new forms of knowledge production (Song, 2025). From this perspective, curriculum evaluation is a core quality-assurance mechanism because it provides evidence regarding how effectively educational intentions are translated into classroom practices and learning opportunities. The distinction between the intended and implemented curriculum is especially important, since well-designed educational policies may not necessarily produce desirable outcomes if teachers lack adequate resources, professional preparation, or contextual support (Fullan, 2016). Research has shown that teachers' interpretation of curricular objectives substantially affects the degree of alignment between formal plans and classroom implementation (Phaeton & Stears, 2016). Moreover, the quality of textbooks and instructional materials is directly related to students' opportunities to learn, making the production stage a critical link between curriculum design and classroom practice (Sievert et al., 2021). Assessment constitutes another essential dimension because contemporary approaches increasingly conceptualize assessment as a formative process that provides feedback for improving teaching and learning rather than merely assigning grades (William, 2011). Professional development can also strengthen teachers' capacity to implement formative

assessment and other curricular innovations effectively (Parmigiani et al., 2025). In addition, broader educational research suggests that curriculum development should respond to students' psychological and behavioral needs and recognize the influence of both formal and hidden curricular experiences on students' development (Mallillin & Mallillin, 2024; Victor Ojorka et al., 2024). In Afghanistan, curriculum evaluation has particular significance because the educational system has experienced prolonged political instability, conflict, structural disruption, repeated policy changes, and uneven access to educational resources. Historical analyses indicate that the development of formal education in Afghanistan has repeatedly been influenced by social and political transformations (Andishmand, 2010), while regional historical experience also demonstrates the close relationship between political change and educational organization (Aghazadeh, 2020). Analyses of Afghanistan's post-2001 national primary curriculum have identified gaps between declared educational intentions and the actual content and implementation of schooling (Amiri, 2016). International reports further indicate continuing challenges related to educational access, instructional resources, learning quality, and regional inequality (Unesco, 2023; Unicef, 2025). More broadly, recent international evidence emphasizes the need for educational systems to monitor learning quality systematically, particularly following periods of disruption and inequality (Oecd, 2023). Therefore, a multidimensional and multi-stakeholder evaluation of curriculum performance is necessary. Teachers and school administrators can provide direct evidence about implementation, classroom feasibility, instructional resources, and assessment, whereas experts and specialists can evaluate design, production, and policy-level coherence. Utilization-focused evaluation supports this type of approach because evidence becomes more valuable when it can directly guide program improvement and decision making (Patton, 2008). The validity of such evidence also depends on the methodological quality of the measurement instruments, including appropriate assessment of content validity, convergent validity, reliability, and factorial structure (Hair, 2014; Polit & Beck, 2006). Therefore, the present study aimed to quantitatively evaluate the performance of primary school curricula in the Afghan education system from the perspectives of teachers, administrators, experts, and specialists, with particular attention to curriculum design and development, production, implementation, evaluation, and proposed improvement strategies.

Methods and Materials

This applied study employed a descriptive survey design with a correlational-comparative approach. The statistical population consisted of teachers, school administrators, educational experts, and specialists working in primary education in six provinces of Afghanistan, namely Kabul, Bamyan, Mazar-e-Sharif, Kandahar, Farah, and Herat, during the 2023–2024 academic year. The teacher population was estimated at approximately 19,800 individuals. Stratified random sampling was conducted by province, and the sample size was determined using the Morgan sampling table. In total, 650 questionnaires were distributed among teachers and administrators and 170 among experts and specialists. Of these, 590 and 160 questionnaires, respectively, were returned. Questionnaires with more than 10% missing responses or uniform response patterns were excluded. The final analytic sample consisted of 572 teachers and administrators and 153 educational experts and specialists, yielding a total of 725 participants. Data collection was conducted over four months through both in-person administration and electronic distribution to schools and provincial education offices. Sporadic missing data representing less than 2% of all responses were replaced using the series mean method in SPSS. Two researcher-developed questionnaires were used. The teacher and administrator questionnaire contained 50 closed-ended items across five subscales: design and development, implementation, production, evaluation, and suggestions and strategies. The expert and specialist questionnaire contained 30 items across three subscales: design and development, production, and suggestions and strategies. Each subscale contained 10 items. Responses were scored on a five-point Likert scale ranging from 1, strongly disagree, to 5, strongly agree, with 3 considered the theoretical midpoint. Higher scores indicated more favorable evaluations. The instruments were developed through a systematic review of the literature and preliminary exploratory interviews with specialists. Before final administration, both questionnaires were pilot-tested with 30 teachers and experts outside the main sample to evaluate clarity, completion time, and item ambiguity. Minor revisions were made to item wording without altering the conceptual structure. Content validity, convergent validity, composite reliability, and second-order confirmatory factor structure were examined and found to be acceptable. Statistical analyses were conducted using SPSS version 26, while SmartPLS version 3 was used to examine the measurement structure. Normality was assessed using the Kolmogorov–Smirnov test and inspection of histograms and Q-Q plots. Homogeneity of variance was examined using Levene's test, and Welch-adjusted *t* statistics were used when equality of variance assumptions were violated. One-sample *t* tests were used to compare curriculum performance scores with the theoretical midpoint of 3. Friedman tests were used to rank curriculum components. Independent-samples *t* tests were used to compare teachers and experts on shared dimensions, and one-way analysis of variance with Scheffé post hoc testing was used to examine differences according to age, work experience, province, and educational level. Statistical significance was set at .05.

Findings

The inferential analyses indicated that overall curriculum performance was evaluated significantly above the theoretical midpoint by both stakeholder groups. Among experts and specialists, the overall curriculum score was significantly higher than 3, $t(152)$

= 25.82, $p < .01$. All three expert-rated components were also significantly above the theoretical midpoint: design and development, $t(152) = 21.52$, $p < .01$, mean difference = 0.52, 95% CI [0.47, 0.57]; production, $t(152) = 11.23$, $p < .01$, mean difference = 0.35, 95% CI [0.28, 0.41]; and suggestions and strategies, $t(152) = 20.66$, $p < .01$, mean difference = 0.85, 95% CI [0.77, 0.93]. Among teachers and administrators, the overall curriculum score was likewise significantly above the theoretical midpoint, $t(418) = 56.20$, $p < .01$, mean difference = 0.72, 95% CI [0.69, 0.74]. Each of the five components was also significantly above 3: design and development, $t(418) = 29.37$, $p < .01$, mean difference = 0.57; implementation, $t(418) = 42.74$, $p < .01$, mean difference = 0.82; production, $t(418) = 25.62$, $p < .01$, mean difference = 0.56; evaluation, $t(418) = 25.43$, $p < .01$, mean difference = 0.57; and suggestions and strategies, $t(418) = 42.56$, $p < .01$, mean difference = 1.06. The Friedman test revealed significant differences in the relative ranking of components among teachers and administrators, $\chi^2(4, N = 419) = 311.81$, $p < .001$. Suggestions and strategies obtained the highest mean rank of 4.01, followed by implementation with a mean rank of 3.38. Design and development, production, and evaluation occupied lower and relatively similar positions, with mean ranks of 2.54, 2.56, and 2.52, respectively. A significant difference in component rankings was also found among experts, $\chi^2 = 114.18$, $p < .001$, with suggestions and strategies ranked highest and production ranked lowest. Independent-samples t tests comparing the two groups on the three common components showed no significant difference for design and development, $t(570) = -0.95$, $p = .34$. However, significant between-group differences were found for production, $t(570) = -3.54$, $p < .01$, and suggestions and strategies, $t(570) = -2.95$, $p < .01$. Experts evaluated the production and development of textbooks more critically than teachers, whereas teachers expressed stronger agreement with proposed reform strategies. Analysis of demographic and professional variables indicated that age, work experience, and province of employment significantly affected evaluations of selected curriculum components, whereas gender and educational level did not produce significant differences.

Discussion and Conclusion

The findings indicate that the primary school curriculum in Afghanistan is perceived as functioning above an average level overall, yet this favorable general assessment conceals important inequalities among different stages of the curriculum cycle. The fact that all major components exceeded the theoretical midpoint suggests that both school-level practitioners and system-level experts recognize meaningful strengths in the existing curriculum. However, the significant Friedman test results demonstrate that curriculum quality is not uniformly distributed across design, production, implementation, evaluation, and reform-oriented dimensions. This distinction is important because a curriculum can appear acceptable in aggregate while still containing specific operational weaknesses that reduce its overall effectiveness. The comparatively low ranking of production is particularly noteworthy. Textbook development and material production represent the stage at which formal curricular intentions are translated into concrete educational resources. Weaknesses in printing quality, organization, visual presentation, availability, or instructional suitability may therefore undermine otherwise sound curricular objectives. The more critical judgments expressed by experts in this domain may reflect their greater familiarity with technical and professional standards for curriculum material development. Teachers, by contrast, may evaluate materials primarily according to their day-to-day usability in classrooms.

The high ranking of suggestions and strategies should not necessarily be interpreted as evidence that this dimension of the existing curriculum is already highly developed. Rather, it likely reflects strong agreement among participants regarding the need for reform. In this sense, high scores on reform-oriented items may indicate an awareness of existing shortcomings and readiness for systematic improvement. The finding that teachers agreed more strongly with proposed reforms than experts may be explained by their direct exposure to practical classroom constraints. Teachers encounter the immediate consequences of insufficient resources, inappropriate content sequencing, assessment limitations, and student diversity, which may increase their perceived urgency for change. At the same time, the absence of a significant difference between teachers and experts in design and development suggests a relatively shared perception of the formal curriculum framework. The divergence appears to emerge more clearly when the curriculum moves from formal planning into material production and practical implementation.

The relatively high position of implementation among teachers and administrators may also reflect teachers' adaptive capacity. In resource-constrained educational settings, teachers often modify instructional practices to compensate for limitations in textbooks, classroom equipment, or institutional support. Such adaptive practices may explain why implementation was evaluated more positively than production or evaluation. Nevertheless, the relatively lower ranking of evaluation suggests that assessment remains a vulnerable component of the curriculum cycle. If assessment mechanisms are insufficiently aligned with curricular objectives or fail to provide meaningful feedback about students' actual learning, the educational system may struggle to identify learning gaps and adjust instruction effectively.

The significant effects of age, work experience, and province further indicate that curriculum evaluation is partly shaped by professional and contextual experience. Older and more experienced participants may assess current conditions by comparing them with earlier stages of Afghanistan's educational development, whereas younger personnel may judge the curriculum according to more recent expectations of instructional quality and modernization. Provincial differences likely reflect unequal access to instructional resources, professional development opportunities, school infrastructure, and administrative support.

These findings suggest that curriculum reform should not be implemented as a uniform national intervention without regard to contextual variation. Instead, national standards should be combined with context-sensitive implementation strategies that respond to the distinct educational conditions of different provinces.

Overall, the findings show that Afghanistan's primary school curriculum possesses a generally acceptable foundation but requires targeted improvement in specific operational domains. The most important priorities appear to include strengthening textbook production and quality-control procedures, improving classroom assessment practices, expanding teacher professional development, and establishing systematic mechanisms for collecting stakeholder feedback. Reform should be approached as an integrated process rather than as isolated revisions to content. Improvements in curriculum design will have limited value if instructional materials, teacher preparation, implementation conditions, and assessment systems are not improved simultaneously. A sustainable curriculum reform strategy should therefore connect national curriculum policy with the realities of schools and classrooms, ensure continuous participation of teachers and experts in curriculum review, and establish periodic evidence-based monitoring of curriculum performance. Such an approach would allow Afghanistan's primary education system to move from episodic curriculum revision toward a more coherent, responsive, and continuously improving model of curriculum governance.

References

- Aghazadeh, A. (2020). *History of education in Iran*. SAMT.
- Amiri, F. (2016). *Curriculum matters: An analysis of primary schools' national curriculum in Afghanistan post 9/11* [Doctoral dissertation, Lehigh University].
- Andishmand, S. A. (2010). *History of education in Afghanistan*. Franklin Publications.
- Benes, S., & Alperin, H. (2025). *The Essentials of Teaching Health Education: Curriculum, Instruction, and Assessment*. Human Kinetics.
- Fathi Vajargah, K. (2023). *Principles and concepts of curriculum planning*. Elm-e Ostadan.
- Fullan, M. (2016). *The new meaning of educational change*. Teachers College Press.
- Hair, J. F. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage.
- Maleki, H. (2019). *Introduction to curriculum planning*. SAMT.
- Mallillin, L. L. D., & Mallillin, M. C. B. (2024). Psychological Assessment and Behavior of Students: An Intervention for School Curriculum and Development Plan. *Ijams*, 4(2), 51-58. <https://doi.org/10.62225/2583049x.2024.4.2.2437>
- Marsh, C. J., & Willis, G. (2007). *Curriculum: Alternative approaches, ongoing issues* (4 ed.). Pearson.
- Mohsenpour, B. (2021). *Foundations of educational planning*. SAMT.
- Oecd. (2023). *PISA 2022 results: Learning during and from disruption*.
- Ornstein, A. C., & Hunkins, F. P. (2018). *Curriculum: Foundations, principles, and issues* (7 ed.). Pearson.
- Parmigiani, D., Nicchia, E., Pario, M., Murgia, E., Radović, S., & Ingersoll, M. (2025). Training in-service subject teachers to implement formative assessment through action research. *Frontiers in Education*.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4 ed.). Sage.
- Phaeton, M. J., & Stears, M. (2016). Exploring the alignment of the intended and implemented curriculum through teachers' interpretation: A case study of A-level biology practical work. *Eurasia Journal of Mathematics, science and technology education*, 13(3), 723-740.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489-497.
- Sievert, H., van den Ham, A. K., & Heinze, A. (2021). The role of textbook quality in first graders' ability to solve quantitative comparisons: A multilevel analysis. *Studies in Educational Evaluation*, 71, 100949.
- Song, E. (2025). Prompt Engineering as a 21st-Century Literacy: A K-12 Curriculum Design and Assessment Framework. *Aies*, 1(2), 32-47. <https://doi.org/10.6914/aiese.010203>
- Taghipour Zahir, A. (2022). *Curriculum planning for primary schools in the third millennium*. Agah Publishing.
- Unesco. (2023). *Global education monitoring report*.
- Unicef. (2025). *Afghanistan education fact sheet*.
- Victor Ojorka, A. K. O. R., Offor, M., & Ojonugwa, D. S. (2024). Assessment of Hidden Curriculum and Personality Development among University Teacher Education Students. *Research Journal of Humanities and Cultural Studies*, 10(4), 108-117. <https://doi.org/10.56201/rjhcs.v10.no4.2024.pg108.117>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.