

Construction of a Standardized Questionnaire to Detect the Pseudo Evaluation in Elementary Schools

Mahdi Ghorbankhan*
Keyvan Salehi**
Ali Moghaddam Zadeh***

Introduction

Distorting evaluation results is one of the most devastating damage that it threatens the dynamism of any system. The main purpose of this study is to design a Pseudo Evaluation Questionnaire (PEQ) for elementary schools. We conduct two studies for the structure of the PEQ. In the first study, we demonstrate the possible causes of pseudo evaluation in elementary school. In the second study, we use exploratory and confirmatory factor analysis methodologies to examine the factor structure of the PEQ.

Method

In order to achieve research goals, we use exploratory mixed method. In the first study, 30 individuals were selected through Purposeful Sampling and in the second study, 714 teachers were selected through cluster sampling from elementary school teachers in Tehran. We also check its reliability and validity in different ways. For reliability, omega, alpha and marginal coefficients were estimated.

Results

Qualitative data led to the identification of 4 main factors using the Colaizzi method. These main factors are: Teacher professional agency, management and planning, evaluation tools and techniques, interpersonal relationships. In the second study, results of exploratory and confirmatory factor analysis confirmed the structure of Study 1, in addition to providing evidence for the PEQ's predictability. In the next section Graded Response

* Ph.D. Student in Assessment and Measurement, University of Tehran, Tehran, Iran.

** Assistant Professor, University of Tehran, Tehran, Iran.

Corresponding Author: Keyvansalehi@ut.ac.ir

***.Assistant Professor, University of Tehran, Thran, Iran.

Model (GRM) used to examine the characteristics of items and data fit. The result of this section was that 2 items were removed due to inappropriateness. After a thorough examination of the items, 35 items remained.

Discussion

Overall, the results indicate that the PEQ measures the factors of pseudo evaluation which can help to identify possible causes of pseudo evaluation in elementary schools. By identifying these factors, its consequences can be prevented as much as possible.

Keywords: pseudo evaluation, test construction, standardized questionnaire, Pseudo Evaluation Questionnaire (PEQ), graded response model.



Author Contributions: Dr. Keyvan Salehi was responsible for leading the overall research process. Mahdi Ghorbankhani was responsible for research plan design, data collection and analysis and all authors discussed the results, reviewed and approved the final version of the manuscript.

Acknowledgments: The authors thank all dear teachers who have helped us in this research.

Conflicts of interest: The authors declare there is no conflict of interest in this article.

Funding: This research is not sponsored by any institution and all costs have been borne by the authors

مجله‌ی علوم تربیتی دانشگاه شهید چمران اهواز
پاییز و زمستان ۱۳۹۹، دوره‌ی ششم، سال ۲۷
شماره‌ی ۲، صص: ۹۱-۱۱۶

تاریخ دریافت مقاله: ۱۳۹۹/۰۶/۳۰
تاریخ پذیرش مقاله: ۱۳۹۹/۱۰/۱۱

مقاله پژوهشی

ساخت پرسشنامه میزان‌شده برای تشخیص ارزشیابی کاذب در مدارس ابتدایی

مهدی قربانخانی*

کیوان صالحی**

علی مقدم‌زاده***

چکیده

تحریف نتایج ارزشیابی از مخرب‌ترین آسیب‌هایی است که پویایی و نشاط هر نظامی را با چالش‌های جدی همراه می‌سازد. هدف اصلی این پژوهش ساخت پرسشنامه ارزشیابی کاذب برای مدارس ابتدایی است. بدین‌منظور مبتنی بر روش پژوهش آمیخته اکتشافی دو مطالعه به انجام رسید که در مطالعه اول از طریق نمونه‌گیری هدفمند از نوع ملاکی، ۳۰ نفر و در مطالعه دوم، از میان معلمان مدارس ابتدایی شهر تهران تعداد ۷۱۴ نفر از طریق نمونه‌گیری خوشه‌ای سه‌مرحله‌ای انتخاب شدند. داده‌های کیفی با استفاده از راهبرد هفت مرحله‌ای کلاسی به شناسایی چهار عامل اصلی منجر گردید. در مطالعه دوم، به تدوین ابزار بر مبنای یافته‌های مطالعه اول و بررسی ویژگی‌های آن با استفاده از تحلیل عاملی اکتشافی و تاییدی پرداخته شد؛ همچنین از مدل پاسخ مدرج سیم جیما برای بررسی ویژگی گویه‌ها، گزینه‌ها و برازندگی داده‌ها استفاده شد. در نهایت، پس از اعمال اصلاحات و حذف موارد نامناسب ابزار تدوین شده، نهایی گردید.

واژه‌های کلیدی: ارزشیابی کاذب، ساخت ابزار، پرسشنامه میزان‌شده، پرسشنامه ارزشیابی کاذب، مدل چند ارزشی.

* دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه تهران، تهران، ایران

Keyvansalehi@ut.ac.ir

** استادیار، دانشگاه تهران، تهران، ایران (نویسنده مسئول)

*** استادیار، دانشگاه تهران، تهران، ایران

مقدمه

امروزه بر اثر پیشرفت سریع علوم و فناوری، مفهوم آموزش و پرورش به کلی دگرگون شده و دامنه آن نیز گسترش یافته است. نقش تعلیم و تربیت جدید آن است که استعدادها و توانایی‌های دانش‌آموزان را شکوفا نموده و آنها را برای شرکت در فعالیت‌های سازنده جامعه پویای بشری آماده سازد، به نحوی که سعادت واقعی افراد و جوامع تأمین گردد. نیل به این اهداف، مستلزم تغییرات مطلوب در رفتار دانش‌آموزان است. بدین معنی که هر یک از مواد برنامه‌های آموزشی در نگرش‌ها، علایق، توانایی‌ها و رفتار دانش‌آموزان تغییرات معینی را به وجود آورند (Ghorbankhani, Salehi & Moghadamzadeh, 2019). زمانی می‌توان تغییرات این‌چنینی را اعمال کرد که به تشخیص بی‌عیب محیط‌های آموزشی و اصول و قواعد حاکم بر آن نائل آمد و کنش‌های آگاهانه‌ای را اتخاذ کرد. مدارس به‌عنوان پایدارترین شکل نهاد آموزش و پرورش همواره در معرض پرسش‌هایی در قبال کیفیت خود بوده‌اند. عوامل متعددی همچون نیروی انسانی متخصص و متعهد، مدیریت آموزشی قوی، ویژگی‌های فراگیران، فرایندهای یادگیری و یاددهی (Habibi, Pardakhtchi, Abolghasemi & Ghahremani, 2013)، قوانین و مقررات مدرسه، جو مدرسه، فرهنگ مدرسه، انتظارات جامعه از مدرسه (Lai & keung, 2011)، خانواده، محیط زندگی، خدمات مدرسه، معلمان، مدیران، برنامه آموزشی (Shahdosti, 2011) و همچنین کیفیت معماری در محیط‌های یادگیری (Khosrojerdi & Mahmoudi, 2014)، در کیفیت مدارس مؤثرند (Mahdiyoun, Pahng & Yarigholi, 2017). تحقق مدرسه‌ی با کیفیت در تمامی ابعاد از اهمیت برخوردار است به‌خصوص در ابعادی که مرتبط با فرآیند یاددهی و یادگیری در مدارس است؛ از جمله‌ی با اهمیت‌ترین عوامل موجود در فرآیند یاددهی و یادگیری مدارس عامل ارزشیابی است که تغییر و تحریف عمدی و یا سهوی در نتایج آن بسیار تأثیرگذار در سرنوشت دانش‌آموزان و عمدتاً جبران‌ناپذیر است.

پرواضح است که شناسایی دقیق و رواند استعدادها، علایق و نگرش‌های دانش‌آموزان منوط به بهره‌گیری صحیح از یک نظام ارزشیابی بی‌عیب و نقص است و پرواضح است در صورت عدم شناسایی صحیح استعدادها و یا اتخاذ تدابیر نامناسب برای پرورش و زمینه‌سازی برای رشد و تعالی دانش‌آموزان می‌تواند پیامدهای ناگوار و جبران‌ناپذیری به همراه داشته باشد. ارزشیابی صحیح، زمینه‌ای را فراهم می‌کند تا برنامه‌ریزان و مجریان نسبت به قوت‌ها و ضعف، اطلاعات و شناخت لازم را برای اصلاح برنامه‌های موجود و طراحی برنامه‌های با

کیفیت‌تر کسب کنند و مناسب‌ترین روش‌ها و تمهیدات را برای پیشبرد اهداف برنامه به کار بندند (Abedi, 2011). همان‌طور که ذکر شد تحریف در نتایج ارزشیابی می‌تواند پیامدهای جبران‌ناپذیری را به همراه داشته باشد به‌زعم استافل‌بیم^۱ ارزشیابی ممکن است به دلیل تمرکز ناکافی ارزیاب در حین ارزشیابی، اشتباهات فنی در استفاده از فنون و ابزارهای ارزیابی، سوءاستفاده از قدرت برای تأمین خواسته‌ها و تمایلات شخصی، اجرای نامناسب فرآیند ارزشیابی، تأخیر در گزارش یافته‌ها و نتایج ارزشیابی، یافته‌های سوگیرانه و مغرضانه، نتیجه‌گیری‌های ناموجه، تفسیرهای ناکافی یا اشتباه برای کاربران و پیشنهادهای ناموجه، کم‌اعتبار و بی‌ارزش شود (Stufflebeam, 2001). هرگام از فرآیند ارزشیابی (همچون جزئیات ارجاع‌دهی، جمع‌آوری داده‌ها، پیش‌نویس گزارش و نتایج) به طور بالقوه می‌تواند طعمه رفتارهای نامناسب و سوء استفاده باشد (Desautels & Jacob, 2012). ارزیاب برای جلوگیری از بروز انواع مشکلات ارزشیابی که می‌تواند ناشی از عدم کاربست استانداردهای ارزشیابی باشد، باید علاوه بر شناخت برنامه مورد نظر از یک الگو و چهارچوب خاصی، برگرفته از استانداردها استفاده نماید (Ghaemipour, Karami & Khandaghi, 2015).

دوره ابتدایی، مهم‌ترین دوره تحصیلی در همه نظام‌های آموزشی دنیاست؛ چرا که پایه اصلی رشد همه‌جانبه فرد در آن شکل گرفته، شخصیت او پایه‌گذاری می‌شود. این مهم، بر اهمیت و ضرورت مراقبت در این دوره بیش از دوره‌های دیگر صحنه می‌گذارد (Ghorbankhani, et al., 2019). در این دوره می‌بایست با تشخیص درست استعدادها و علایق دانش‌آموزان از طریق سازوکاری منظم و از پیش تعیین‌شده آن‌ها را در مسیری که شایسته است هدایت نمود و با انتخاب اقدامات مقتضی به رشد و شکوفایی استعدادها و پرورش خلاقیت آن‌ها پرداخت. پرواضح است اگر یک نظام آموزشی قادر به تشخیص استعدادها و علایق فردی نباشد، در رسیدن به اهداف پیش‌بینی‌شده موفق نخواهد بود. با توجه به اینکه نظام‌های آموزشی وظیفه بسیار سنگینی را بر عهده دارند، لازم است در طراحی و اجرای فعالیت‌های آن از مطلوب‌ترین شیوه‌ها استفاده کرد (Bazargan, 2012). در واقع تغییرات نظام سنجش و ارزشیابی باید در راستای تغییر بنیادین نظام آموزشی باشد. بهره‌مندی از رویکردهای نوین سنجش آموخته‌ها، در همه‌ی دوره‌های آموزشی ضرورت و اهمیت بسیاری دارد، اما این ضرورت در دوره ابتدایی

1- Stufflebeam

به دلیل حساسیت و جایگاه ویژه‌ی آن در کسب مهارت‌های پایه دوجندان است (Lee, Oh, 2007; Ang & Lee, 2014; Samuel, 2008; Nxumalo, 2007).

(Mohammadzadeh and Salehi, 2017) با استفاده از روش نظریه برخاسته از داده‌ها^۱ به تبیین فرآیند ارزشیابی پیشرفت تحصیلی از منظر قرآن کریم پرداختند و اهم اصول اساسی اجرای صحیح ارزشیابی پیشرفت تحصیلی و تصریح شده در قرآن کریم را اینگونه معرفی می‌نمایند: ارزشیابی باید مبتنی بر این باور باشد که انسان موجودی آزاد و مخیر است؛ انتظارات و کمال مطلوب مدنظر ارزشیابی باید مشخص باشد (استاندارد سنجش و ارزیابی)؛ آغاز و پایان ارزشیابی باید مشخص باشد؛ ارزشیابی باید مکتوب، رسمی و مستند باشد؛ ارزشیابی کیفی و توصیفی همه اعمال در قالب یک مجموعه مکتوب؛ اسناد و گزارش‌های تهیه‌شده باید دقیق، جامع و منصفانه باشد. با توجه به اهمیت ارزیابی در نظام‌های آموزشی و به خصوص نظام آموزش ابتدایی به بحث گذاردن تجربه‌های نامناسب و سؤال برانگیز از ارزشیابی‌ها در مقابل نمونه‌های مطلوب و موفق آن، ضروری به نظر می‌رسد و در حالت کلی اهمیت آموزش و ارزشیابی در دوره ابتدایی را نمی‌توان نادیده گرفت و یا آن را کم‌اهمیت تلقی کرد (Dewhirst, 2019)؛ با توجه به اهمیت این امر ضروری است تا سازوکاری ایجاد نمود که از طریق آن بتوان به شناسایی کیفیت این پدیده در مدارس پرداخت. تعبیر استافل بیم از عواملی که منجر به تحریف نتایج ارزشیابی در هر سطح و سازمانی می‌شوند عوامل تحریف‌کننده^۲ و واژه‌ای که به این نوع از ارزشیابی اطلاق می‌کند ارزشیابی کاذب^۳ است. در واقع ارزشیابی کاذب به آن نوع از ارزشیابی گفته می‌شود که نتایج آن دستخوش تغییرات عمدی و یا سهوی شده باشد؛ در این میان اگرچه فعالیت‌هایی که با نام ارزشیابی انجام می‌شوند ممکن است یک ارزشیابی بی‌عیب و نقص محسوب گردند ولی اگر قادر به انجام ارزشیابی معتبر و همسو با استانداردهای ارزشیابی نباشند، ارزشیابی کاذب نامیده می‌شوند (Shinkfield & Stufflebeam, 2012). شکل‌گیری فرهنگ ارزشیابی کاذب از مخرب‌ترین آسیب‌هایی است که پویایی و نشاط هر نظامی را با چالش‌های جدی همراه می‌سازد. این امر ضرورت آگاهی برنامه‌ریزان و مدیران ارشد از عواملی که زمینه‌ساز یا مشوق ارزشیابی کاذب می‌گردند را دوجندان می‌نماید. اهمیت نتایج ارزشیابی از آن روست که با

1- grounded theory
2- distracting factors
3- psude evaluation

بحث و مذاکره می‌تواند بر تصمیم‌گیری و شکل‌گیری سیاست‌های مختلف هر سازمانی مؤثر باشد (Sonpal-Valias, 2009; Valovirta, 2002) البته باید توجه داشت که بدون عنایت به روش‌های علمی ارزشیابی، هر ارزشیابی منتج به هدف‌های مطلوب نخواهد شد و به طور قطع، نیروهای مادی و انسانی را به طور نسبی به هدر خواهد داد.

با بررسی‌های صورت گرفته در پیشینه‌ی این موضوع پژوهش‌های متعددی به بررسی ارزشیابی و روش‌های انجام صحیح آن پرداخته‌اند که در ادامه به برخی از آن‌ها اشاره می‌شود در این میان تنها یک مطالعه (Stufflebeam & Shinkfield, 2012) به بررسی عوامل تحریف‌کننده نتایج ارزشیابی پرداخته و از آن با عنوان ارزشیابی کاذب یاد کرده است.

به باور استافل بیم ارزشیابی کاذب زمانی رخ می‌دهد که ارزیاب خواسته یا ناخواسته در گزارش خود آنچه که با واقعیت مطابق نیست را به سازمان سفارش‌دهنده ارزشیابی ارائه و یا به صورت عمومی و تصمیم‌گیری در آن مورد منتشر کند. ممکن است گاهی اوقات ارزیابان به دلایل مختلفی همچون سوگیری^۱، انگیزه منفعت‌طلبی^۲، فریب‌خوردن^۳ و اعمال سلیقه^۴ به صورت‌های مختلفی مثل اعلام‌گزینشی^۵، تعمیم کلی^۶، مخفی‌کاری^۷، تحریف و دستکاری یافته‌ها^۸، تأمین آموزش‌های ارزیابی^۹، توسعه توانمندی ارزیابی^{۱۰} و دریافت و ارائه بازخورد^{۱۱}، واقعیت‌ها را آنگونه که هستند، بازنمایی نکنند (Stufflebeam & Shinkfield, 2012).

در منابع مختلف از ویژگی‌های ارزشیابی نامناسب که مبتنی بر شیوه‌های سستی ارزشیابی و بدون توجه به کاربست آموخته‌ها در محیط‌های واقعی زندگی است به مواردی همچون: عدم بهره‌گیری برای اصلاح روش‌های آموزش (Seif, 2012) عدم بهره‌گیری برای رفع اشکالات یادگیری دانش‌آموزان (Shabani, 2017) سروکار داشتن با حافظه به جای ظرفیت‌های خلاق دانش‌آموزان و دخیل بودن عامل شانس و تصادف در نتیجه ارزشیابی (Shoarinejad, 2015)

- 1- bias
- 2- profit motive
- 3- deceived
- 4- evaluators perspective
- 5- selectively release
- 6- overgeneralize
- 7- hiding
- 8- falsify findings
- 9- providing evaluation training
- 10- providing an organizations evaluation capability
- 11- providing feedback

نمره‌گرا بودن (Sarmad, 2010) عدم توانایی در سنجش میزان تغییر رفتارهای آزمودنی در بلندمدت (Porzahir, 2016; Ghorbankhani & Salehi, 2017) اشاره شده است. با توجه اهمیت دوره ابتدایی در یک نظام آموزشی، شناسایی محدوده ارزیابی معتبر و مورد تأیید استانداردهای ارزشیابی و همچنین به تبع آن تعیین و شناسایی محدوده آنچه که به عنوان ارزشیابی کاذب بدان اشاره شد، در این دوره ضروری است. با توجه به مفهوم‌پردازی‌های صورت‌گرفته در خصوص رواج ارزشیابی کاذب در مدارس که اخیراً در ادبیات علمی این حوزه به همت برجستگان حوزه ارزشیابی همچون استافل بیم و شینکفیلد ضروری به نظر می‌رسد تا بر این ردپای این پدیده را در مدارس ابتدایی کشورمان با ابزاری روا و پایا شناسایی و از خطرات و پیامدهای ناگوار آن پیشگیری کرد؛ بر این اساس هدف کلی پژوهش، ساخت پرسشنامه میزان‌شده تشخیص ارزشیابی کاذب در مدارس ابتدایی است و با توجه به این هدف، پرسش‌های پژوهش عبارتند از:

۱- بر اساس تجربه‌زیسته مطلعان، چه عواملی در مدارس ابتدایی، زمینه‌ساز تحریف نتایج ارزشیابی می‌شوند؟ ۲- آیا پرسشنامه طراحی‌شده برای تشخیص ارزشیابی کاذب در مدارس ابتدایی، از ویژگی‌های روان‌سنجی مناسبی برخوردار است؟

روش

با توجه به اینکه پاسخ‌دهی به مسئله پژوهش حاضر در برگزیده دو بخش الف) شناسایی مؤلفه‌های ارزشیابی که قابلیت تحریف دارند و ب) تعیین ویژگی‌های روان‌سنجی ابزار طراحی شده است، پژوهش حاضر به روش آمیخته اکتشافی انجام شده است. به باور Teddlie and Tashakori (2003) ترکیب دو روش کمی و کیفی، زمینه‌ی مناسب برای یکپارچگی فنون جمع‌آوری و تحلیل داده‌های هر دو روش کمی و کیفی در مراحل همزمان یا زنجیره‌ای را مهیا ساخته و در مقایسه با به‌کارگیری هر یک از این دو روش به‌تنهایی به درک بهتری از مسائل پژوهشی منجر می‌شود (as cited in Salehi & Golafshani, 2010). بخش اول این پژوهش مطالعه‌ای کیفی و از نوع پدیدارشناسی است که به منظور شناسایی مؤلفه‌های ارزشیابی که به طور بالقوه قابلیت تحریف دارند، انجام گردید. تلاش شد تا به صورت هدفمند و با راهبرد نمونه‌گیری ملاکی با مدیران مدارس و آموزگاران واجد شرایط (داشتن سابقه تدریس بالای ده سال و همچنین داشتن تجربه تدریس در مدارس غیر دولتی و دولتی)، به صورت نیمه‌ساختاریافته مصاحبه شود و در مصاحبه ۲۷ اشباع نظری حاصل شد و ۳ مصاحبه برای اعتباربخشی شد.

برای ارتقای اعتبارپذیری یافته‌ها (معادل روایی درونی در پژوهش کمی)، از فنون و راهبردهایی نظیر حضور طولانی مدت در محیط پژوهش و صرف زمان طولانی با شرکت کنندگان، تبادل نظر با هم‌تایان، کنترل از سوی اعضا، استفاده شد. همچنین از سه‌سویه‌نگری مبتنی بر تنوع بخشی در تعداد سؤال‌ها، تحلیل موارد منفی و متناقض تفسیر شده در داده‌ها، و نظرات چهار نفر از استادان و متخصصان تعلیم و تربیت، به منظور اعمال نظر نهایی، بهره‌گرفته شد. همچنین تلاش شد تا با توصیف دقیق زمینه مطالعه، پیش‌فرض‌ها و شرایط پژوهش و ارائه جزئیات آن، زمینه لازم برای افزایش انتقال‌پذیری یافته‌های پژوهش (معادل روایی بیرونی در پژوهش کمی) فراهم شود. به‌منظور ایجاد قابلیت اطمینان به پژوهش (معادل‌پایایی در پژوهش کمی)، تلاش شد تا با تشریح جزئیات مربوط به چگونگی گردآوری داده‌ها، نحوه تصمیم‌گیری‌ها، تفسیر و تحلیل‌های طی‌شده در فرایند پژوهش، زمینه‌ایی برای انجام «مسیرنمای حسابرسی» فراهم شود. همچنین برای ارتقای تاییدپذیری یافته‌های پژوهش (معادل عینیت در پژوهش کمی)، تلاش شد تا با استفاده از یادداشت میدانی و در فرایند تحلیل و تفسیرهای انجام‌شده، تا آن‌جا که امکان‌پذیر است، اندیشه‌ها، باورها و انتظارات ذهنی پژوهشگران در جریان گردآوری و تحلیل داده‌ها کنترل شود (Ghorbankhani & Salehi, 2015). به باور Abolmaali (2012)، با توجه به اینکه نگارش تحلیل داده‌ها در روش کیفی، متضمن گزارش داده‌های متنی است و با لغات، موضوعات و نوشته‌ها مشخص می‌شود، اغلب از نقل قول‌های مستقیم شرکت‌کنندگان و از ضمیر اول شخص مفرد، یا در اصطلاح «صدای فعال» استفاده می‌شود (as cited in Salehi, Bazargan, Sadeghi & Shokohi-yekta, 2015). چنین صدایی منجر به اطمینان بیشتر به یافته‌های پژوهش کیفی می‌شود. در پژوهش حاضر سعی کردیم تا این ویژگی‌ها در نگارش بخش کیفی روش و یافته‌های مربوطه به آن مورد توجه ویژه قرار گیرد.

برای تحلیل داده‌ها اقداماتی نظیر مرور یادداشت‌های تدوین‌شده بلافاصله بعد از انجام هر مصاحبه برای تعدیل و تکمیل مطالب یادداشت شده و همچنین تحلیل با استفاده از راهبرد هفت مرحله‌ای گلازی و نرم افزار مکس.کیودی^۱ انجام شد. بدین منظور گام‌های زیر برداشته شد:

۱. برای آشنایی با عمق و گستره‌ی داده‌ها، روایت‌های ارائه شده که به طور مرسوم پروتکل نام دارد، چندین مرتبه مطالعه کردم تا با آنها مانوس شوم. بدین‌منظور از طریق راهبردهایی

1- active voice

2- Maxqda

نظیر بازخوانی مکرر داده‌ها، خواندن داده‌ها به صورت فعال و جستجوی معانی و الگوها، تلاش کردم تا آشنایی قابل توجهی با داده‌های گردآوری شده فراهم شود؛ ۲. به منظور استخراج جملات مهم، به هر پروتکل مراجعه کردم، جملات و عباراتی که مستقیماً با موضوع مورد مطالعه مرتبط بود را استخراج نمودم؛ ۳. برای فرموله کردن معانی، ضمن مرور منظم و مستمر جملات استخراج شده، تلاش کردم تا به درک معانی هر یک از جملات، پی ببرم؛ ۴. مراحل فوق را برای تمامی مصاحبه‌ها تکرار کردم و معانی فرموله شده و مرتبط به هم را در قالب خوشه‌هایی از مضامین (خوشه‌های معنایی) قرار دادم؛ ۵. در ادامه نتایج را در قالب توصیفی جامع از موضوع مورد مطالعه، تلفیق نمودم؛ ۶. به منظور درک ساختار ذاتی پدیده، تلاش کردم تا توصیف جامعی از پدیده مورد مطالعه را به صورت بیانیه‌ای صریح و روشن از ساختار اساسی پدیده، تدوین نمایم؛ ۷. برای ارتقای پایایی یافته‌ها از شیوه‌ی بازرسی و بازبینی در زمان کدگذاری، استفاده از نظرات همکاران پژوهشی و تأیید آن‌ها (ممیز بیرونی) و درگیری طولانی مدت بهره‌گیری شد. در شیوه‌ی بازرسی و بازبینی در زمان کدگذاری به دست آمده، برای جلوگیری از بدفهمی، اطلاعات کسب شده را دوباره به اطلاع رسانی‌ها نشان دادم تا در صورت نیاز به تعدیل، اصلاحات لازم را انجام داده و تأییدیه نهایی آن‌ها را اخذ نمایم. در روش استفاده از نظرات همکاران پژوهشی و تأیید آن‌ها، نویسنده دوم که به طور مستمر بر روش کار و فرایند تحلیل‌گر نظارت داشت، روند کار را مورد بازبینی قرار داده و اصلاحات مورد نیاز را به تحلیل‌گر ارائه می‌داد. همچنین در برخی از موارد مورد نیاز، از نظرات برخی دیگر از استادان بهره گرفته شد. علاوه بر این، محققین بیش از هشت ماه مستمر و منظم درگیر اجرای پژوهش بودند که بخشی از این زمان به حضور نویسنده اول در مدارس و تعامل با معلمان و سایر عناصر انسانی در فضای مدرسه اختصاص یافت.

در مطالعه دوم، برای ساخت و تحلیل ویژگی‌های روان‌سنجی ابزار ساخته شده، از روش همبستگی استفاده شد. جامعه آماری در بخش کمی، شامل همه معلمان مدارس ابتدایی شهر تهران در سال ۹۷-۱۳۹۶ بود. در مجموع ۷۱۴ نفر از معلمان مورد مطالعه قرار گرفت. روش نمونه‌گیری خوشه‌ای سه مرحله‌ای بود به این شکل که ابتدا مناطق آموزشی تهران به پنج بخش جغرافیایی (شمال، شرق، غرب، مرکز و جنوب) تقسیم و سپس از هر بخش دو منطقه‌ی آموزشی به صورت تصادفی انتخاب شد و در هر منطقه دو مدرسه انتخاب شده و پرسشنامه مقدماتی در این مدارس اجرا شد. نسخه اولیه پرسشنامه مشتمل بر ۵۵ گویه بود که به صورت طیف لیکرت

و در قالب یک پیوستار از ۱ تا ۴ به گویه‌ها نمره داده می‌شد. روایی پرسشنامه طی چند مرحله توسط گروهی از متخصصان موضوعی و استادان راهنما و مشاور مورد بررسی و پالایش قرار گرفت و از میان گویه‌های اولیه آنهایی که بیشترین ارتباط با سازه مورد نظر داشتند انتخاب شدند (Tavakoli, Ghanbaripناه, Ghorbankhani & Zivari, 2019). پس از طراحی نسخه اولیه پرسشنامه، به منظور بررسی میزان روایی صوری پرسشنامه و گویه‌های آن، نسخه اولیه ابزار در اختیار استادان، دانشجویان دکتری و خبرگان موضوعی قرار گرفت و از آن‌ها خواسته شد درباره میزان تناسب هریک از گویه‌ها اظهار نظر کنند؛ پاسخ‌دهندگان نیز با توجه به طیف ارائه‌شده میزان تناسب هر گویه را علامت‌گذاری کرده و در مواردی که هر یک از گویه‌ها نیازمند اصلاح بود، در محل پیش‌بینی شده و یا بر روی گویه مورد نظر پیشنهادهای خود را بیان داشتند. پس از گردآوری نسخه‌های در اختیار اظهار نظرکنندگان، تک تک فرم‌ها توسط تیم پژوهش مورد بررسی و پالایش قرار گرفت و گویه‌های نامناسب حذف شد؛ در ادامه به منظور بررسی روایی محتوایی ابزار پنج تن از متخصصان موضوعی که در این حوزه از استادان برجسته و صاحب نام هستند انتخاب شده و از ایشان برای همکاری دعوت به عمل آمد؛ داوران نظرات خود را بر روی طیف چهارگزینه‌ای در خصوص میزان انطباق محتوا و گویه‌ها ابراز داشتند؛ پس از گردآوری فرم‌های پر شده توسط داوران ضریب روایی محتوایی ابزار محاسبه شد؛ برای بررسی روایی سازه از تحلیل عاملی استفاده شده است که در بخش یافته‌ها به جزئیات آن اشاره شده است و همچنین به منظور بررسی پایایی عامل‌ها از شاخص امگا^۱ استفاده شد. در نهایت نسخه بازنگری‌شده پرسشنامه در فضای نمونه اجرا و برازش گویه‌ها با مدل پاسخ مدرج (GR) نظریه سؤال پاسخ بررسی شد که در انتها پرسشنامه نهایی با ۳۵ گویه تدوین شد.

یافته‌ها

همان‌گونه که ذکر شد پژوهش حاضر مبتنی بر رویکرد آمیخته است و بر همین اساس نیز یافته‌ها در دو بخش با عناوین؛ یافته‌های مطالعه اول (بخش کیفی) و یافته‌های مطالعه دوم (بخش کمی) گزارش شده است.

یافته‌های مطالعه اول

سؤال اول: بر اساس تجربه‌زیسته مطلعان، چه عواملی در مدارس ابتدایی، زمینه‌ساز

تحریف نتایج ارزشیابی می‌شوند؟

به منظور پاسخ به این سؤال، با ۳۰ نفر از مدیران و معلمان با تجربه مصاحبه شد و سپس مصاحبه‌ها با استفاده از راهبرد هفت مرحله‌ای گلایزی، تحلیل گردید؛ داده‌های حاصل از ضبط و یادداشت‌برداری‌های به عمل آمده از مصاحبه‌ها، خط به خط خوانده شده و سپس مفاهیم جملات اصلی استخراج و ملاک‌ها و عوامل تشکیل و در نهایت دسته‌بندی شدند. پس از کدگذاری، طی چندین مرحله، کدهای تهیه شده از مصاحبه‌ها مورد بازنگری پژوهشگر قرار گرفت و پس از اعمال اصلاحات مورد نظر به تأیید رسید. بر اساس یافته‌های کیفی، پاسخ سؤال اول پژوهش به دست آمده یعنی عواملی که به طور بالقوه می‌توانند نتایج ارزشیابی را تحریف کنند مورد شناسایی قرار گرفت. پس از شناسایی نشانگرها، با توجه معانی و متن مصاحبه‌ها دسته‌بندی شدند و خوشه‌های معنایی تشکیل شدند که یافته‌های آن در جدول ۱ ارائه شده است.

جدول ۱. عوامل و نشانگرهای زمینه‌ساز در تحریف نتایج ارزشیابی در مدارس ابتدایی

Table 1. Factors and indicators underlying the distortion of evaluation results in elementary schools

عامل	نشانگر	Indicator
عاملیت حرفه‌ای افت انگیزه، افت تمرکز به دلیل حجم بالای فعالیت‌ها؛ سوگیری‌های نگرشی؛ شناخت معلمان	قبل از دانش آموز و خانواده‌ها توسط معلم؛ عملکرد معلمان پایه‌های قبلی...	شناخت معلمان
برنامه‌ریزی و ساختاری	تدوین ناکارآمد برنامه درسی؛ انتخاب محتوا درسی فراتر از توانمندی معلمان؛ ابهام در مفاد آیین نامه‌ها؛ نداشتن راهنمای عملکرد گام به گام آیین نامه‌ها ...	برنامه‌ریزی و ساختاری
امکانات و ابزار	انتخاب ابزار نامناسب برای ارزشیابی؛ اجرایی نبودن ملاک‌ها برای جمع‌بندی عملکرد دانش‌آموزان...	امکانات و ابزار
تعاملات بین فردی	همسو نبودن جامعه فرهنگی با تحولات حوزه آموزش و ارزشیابی؛ روابط و تعاملات (دوستانه یا خصمانه) بین معلمان با مدیران مدرسه؛ رقابت معلمان هم‌پایه در یک مدرسه؛	تعاملات بین فردی

یافته‌های مطالعه دوم

سؤال دوم: آیا پرسشنامه طراحی شده برای ارزشیابی کاذب در مدارس ابتدایی، از ویژگی‌های روان‌سنجی مناسبی برخوردار است؟ به منظور پاسخ‌دهی به این سؤال پژوهش اقدامات ذیل انجام شد: ۱- بررسی روایی ابزار ۲- بررسی اعتبار و ۳- بررسی ضرایب تمیز و دشواری سؤال‌ها (بر اساس نظریه‌های کلاسیک و سؤال-پاسخ).

بررسی روایی

پس از بررسی روایی صورتی، به شرحی که در بخش روش به آن اشاره شد، به منظور

بررسی روایی محتوایی ابزار، ۵ تن از متخصصان موضوعی که در این حوزه از اساتید برجسته و صاحب نام هستند حاضر به همکاری شدند و آزمون را از برای انطباق محتوا و گویه‌ها مورد بررسی قرار دادند. از ایشان خواسته شد تا مرتبط بودن هرگویه را بر اساس یک طیف ۴ گزینه‌ای مشخص نمایند که مربوط بودن هر گویه را به صورت ۴ کاملاً مربوط است؛ ۳ مربوط است؛ ۲ نسبتاً مربوط است و ۱ مربوط نیست نشان می‌داد؛ براساس روش والتز و بازل چنانچه شاخص روایی محتوا بالاتر از ۰/۷۹ باشد مورد تایید است و گویه‌هایی که کمتر از این مقدار بودند کنار گذاشته شده و همه گویه‌های باقی مانده بالاتر از ملاک تعیین شده بود.

پیش از اجرای تحلیل، داده‌ها به لحاظ دارا بودن شرایط مناسب مورد بررسی قرار گرفت و با داده‌های گم‌شده^۱ و پرت^۲ برخورد لازم انجام شد و سپس به بررسی مفروضات پرداخته شد که شامل خطی بودن^۳، نرمال بودن چندمتغیری^۴، یکسانی پراکندگی^۵ خطاها^۶ بود که همگی مجوز انجام ادامه تحلیل را به پژوهشگران می‌داد؛ سپس داده‌ها به دو بخش تقسیم شد که بخش اول برای شناسایی ساختار و بخش دوم برای تایید ساختار مورد استفاده قرار گرفت. همچنین وجود شرایط مناسب برای انجام تحلیل عاملی نیز مورد بررسی قرار گرفت که شامل آزمون بارتلت^۶، شاخص کی، ام، او^۷ و دترمینان بود که همگی اجازه قابلیت اجرای تحلیل عاملی را تایید می‌کرد؛ برای تعیین تعداد عامل‌ها شاخص مقادیر ویژه^۸ و نمودار مبتنی بر آن‌ها و تحلیل موازی^۹ مورد توجه بود که ۴ عامل در همه روش‌ها استخراج شد و ارزش ویژه عامل‌ها به ترتیب ۲۶/۰۱؛ ۵/۵۴؛ ۴/۰۴ و ۲/۲۲ بود. همه عوامل ۵۵ درصد از واریانس کل را تبیین می‌کرد. جدول ۲ گویه‌های دارای همبستگی معنی‌دار با عامل‌ها را نشان می‌دهد که ۲۰ گویه به دلیل بار عاملی پایین و غیرمعنی‌دار از روند تحلیل کنار گذاشته شدند. پس از آنکه عامل‌ها شناسایی شدند لازم است تا اعتبار عامل‌ها مورد بررسی قرار بگیرد که از امگا استفاده شده است؛ در این روش برای سهولت اجرا گویه‌های هر عامل را به صورت یک خرده آزمون

-
- 1- missing
 - 2- outlier
 - 3- linearity
 - 4- multivariate normality
 - 5- homosdasticity
 - 6- Bartlett's test
 - 7- KMO
 - 8- eigenvalue
 - 9- parallel analysis

جدول ۲. بارهای عاملی گویه‌ها
Table 2. Factor loads of items

عوامل Factors				گویه Item	عوامل Factors				گویه Item
4	3	2	1		4	3	2	1	
	.640			Q12				.582	Q1
	.764			Q17				.697	Q2
	.720			Q20				.753	Q8
	.708			Q19				.758	Q9
	.718			Q21				.743	Q14
	.707			Q10				.567	Q16
	.770			Q30				.442	Q22
.893				Q3				.456	Q33
.897				Q13				.556	Q34
.695				Q23				.642	Q35
.640				Q24				.691	Q5
.732				Q29		.694			Q4
.850				Q32		.836			Q7
.842				Q31		.642			Q11
.828				Q6		.997			Q15
						.543			Q25
						.644			Q27
						.776			Q28

تعریف کرده و برای هرکدام به صورت جداگانه در R برآورد شد که البته باید توجه داشت برای برآورد این شاخص نباید داده‌های گمشده داشت. برای نام‌گذاری گویه‌ها از مصاحبه‌های مرحله قبل کمک گرفته شد.

جدول ۳. ماتریس همبستگی عوامل
Table 3. Factor correlation matrix

4	3	2	1	
			1	عاملیت حرفه‌ای معلم
		1	.53*	برنامه‌ریزی و ساختاری
	1	.49*	.59*	امکانات و ابزار
1	.64*	.53*	.63*	تعاملات بین فردی

*رابطه در سطح ۰۰۱. معنی‌دار است.

جدول ۳ روابط میان ابعاد پرسشنامه ارزشیابی کاذب را به عنوان شاخصی از روایی همگرا

نشان می‌دهد؛ همانگونه که از داده‌های جدول برمی‌آید ابعاد پرسشنامه با یکدیگر همبستگی قابل قبول و معنی‌داری را به همراه دارند.

در گام بعدی به بررسی ساختار عاملی شناسایی‌شده در گام قبل پرداخته شد که شامل بررسی مسیرهای رسم شده از عامل‌ها به نشانگرها و بررسی برازش کلی مدل بود.

جدول ۴. ضرایب عاملی استاندارد Table 4. Standard factor coefficients

گویه Item	ضرایب استاندارد Standard coefficients	گویه Item	ضرایب استاندارد Standard coefficients	گویه Item	ضرایب استاندارد Standard coefficients	گویه Item	ضرایب استاندارد Standard coefficients
Q1	.77	Q20	.69	Q5	.72	Q32	.71
Q2	.80	Q19	.79	Q4	.74	Q31	.80
Q8	.81	Q21	.72	Q7	.70	Q6	.77
Q9	.70	Q10	.73	Q11	.74		
Q14	.80	Q30	.68	Q15	.67		
Q16	.64	Q3	.76	Q25	.69		
Q22	.76	Q13	.70	Q27	.76		
Q33	.66	Q23	.77	Q28	.77		
Q34	.70	Q24	.65	Q12	.73		
Q35	.81	Q29	.69	Q17	.66		

جدول ۵. شاخص‌های نیکویی برازش Table 5. Goodness of fit indices

شاخص‌ها Indices	برآورد Estimate
نسبت مجذور خبی به درجه آزادی (CMIN/DF)	2.133
شاخص ریشه دوم میانگین مربعات باقی مانده (RMSEA)	0.029
شاخص نکویی برازش (GFI)	.965
شاخص تعدیل شده نکویی برازش (AGFI)	.939
شاخص برازش تطبیقی (CFI)	.920
شاخص برازش هنجار شده بنتلر بونت (NFI)	.930
شاخص برازش افزایشی (IFI)	.930
شاخص برازش نسبی (RFI)	.923

با توجه به مندرجات جدول ۵، مقادیر CMIN/DF ۲/۱۳۳ و مقادیر RMSEA برابر با ۰/۰۲۹ برآورد شده است و همچنین میزان نکویی برازش (GFI)، شاخص تعدیل شده نکویی برازش (AGFI)، شاخص برازندگی تطبیقی (CFI)، شاخص برازش هنجار شده بنتلر بونت (NNFI)،

شاخص برازش افزایشی (IFI) و شاخص برازش نسبی (RFI) نیز در محدوده بالاتر از ۰/۹ برآورد شده‌اند.

با توجه به شاخص‌های برازش جدول ۵ می‌توان نتیجه گرفت که ابزار اندازه‌گیری ارزشیابی کاذب در مدارس ابتدایی از برازش مطلوبی برخوردار است و با توجه به مقادیر شاخص‌های برازش الگوی نهایی و مرز مقادیر قابل قبول که در بالا ذکر شد، می‌توان گفت که مدل ارائه شده در این تحقیق مورد قبول است. ضرایب رگرسیونی مدل نشان می‌دهد که متغیرهای ذکر شده بخوبی ارزشیابی کاذب را تبیین می‌کند. به طور کلی برازش مؤلفه‌های به دست آمده نشان می‌دهد مؤلفه‌ها دارای بار عاملی مناسبی بوده و معنی‌داری آنها کمتر از ۰/۰۵ است که نشان می‌دهد عوامل هویت حرفه‌ای معلمان، مدیریت و برنامه‌ریزی، ابزارها و فنون ارزشیابی و روابط بین فردی، عوامل مؤثر در بروز ارزشیابی کاذب هستند.

بررسی اعتبار

در این پژوهش ضرایب پایایی برای مؤلفه‌ها و کل ابزار گزارش شده و در راستای ویژگی‌های کلاسیک اندازه‌گیری دسته‌بندی شده است. جدول شماره ۷ ضریب پایایی ابزار را برای هر عامل و کل ابزار گزارش می‌کند. همچنان که از نتایج جدول مشخص است هیچ ضریبی زیر ۰/۷ نیست، ابزار به لحاظ برخوردار از ثبات و دقت در اندازه‌گیری در وضعیت مطلوب همگونی درونی قرار دارد.

جدول ۷. ضرایب پایایی ابزار

Marginal	حاشیه‌ای	Alpha	آلفا	Omega	امگا	Factor	عامل
.899		.904		0.89		عاملیت حرفه‌ای معلم	
.915		.917		0.90		عامل مرتبط با برنامه‌ریزی	
.795		.745		0.78		عامل مرتبط با ابزار	
.789		.797		0.85		عامل مرتبط با روابط بین فردی	
.949		.919		0.89		کل ابزار	

بررسی ضرایب تمیز و دشواری سؤالات

همه ویژگی‌های آزمون را می‌توان از روی ویژگی تک تک سؤالات و یا رابطه بین سؤالات آزمون معین کرد، متداول‌ترین موارد استفاده از تحلیل سؤال، انتخاب سؤالات بهتر

و مناسب‌تر برای تشکیل فرم نهائی آزمون است تا سؤال‌های ضعیف عملکرد آزمون را مخدوش نکند (Rabaniparsa, Mesrabadi & Yarmohammadzade, 2019). برای بررسی این امر از ضریب دشواری و تمیز سؤال‌ها استفاده شد که نتایج آن در جدول ۸ آمده است؛ نتایج به‌دست آمده گویای آن است که در نظریه کلاسیک ضریب‌دشواری اکثر سؤال‌ها در سطح

جدول ۸. پارامترهای هر گویه بر اساس نظریه‌های کلاسیک و سؤال پاسخ (مدل پاسخ مدرج)

Table 8. Parameters of each item based on CTT and IRT (Graded Response Model)

گویه Item	نظریه سؤال پاسخ (IRT)			نظریه کلاسیک (CTT)		
	B3	B2	B1	پارامتر تشخیص (A)	دشواری Difficulty	تمیز Determination
1	1.52	.05	-1.06	1.55	.545	.582
2	1.31	.00	-.94	2.58	.678	.631
3	1.41	-.16	-1.38	1.84	.626	.542
4	1.23	.06	-.98	2.29	.683	.566
5	1.40	.14	-1.02	1.92	.586	.559
6	1.29	.24	-.79	2.71	.661	.611
7	1.23	-.01	-.94	2.63	.663	.579
8	1.39	.08	-.79	2.32	.709	.640
9	1.49	.30	-.88	1.92	.582	.472
10	1.51	-.23	-1.09	1.58	.644	.304
11	1.59	.49	-.56	2.50	.662	.570
12	1.42	.36	-.72	2.43	.708	.605
13	1.80	.40	-.77	1.92	.581	.527
14	1.33	.11	-.78	2.56	.678	.598
15	1.54	.19	-.81	1.90	.584	.551
16	1.41	.26	-.69	2.63	.647	.570
17	1.30	.36	-.69	2.69	.671	.540
18	1.22	.32	-.50	2.83	.677	.558
19	1.92	-.33	-1.49	1.19	.511	.284
20	.90	-.57	-1.42	2.35	.635	.293
21	2.20	-2.36	-5.15	1.55	.367	.148
22	1.01	-.44	-1.46	2.39	.569	.209
23	1.43	-.66	-1.56	2.04	.680	.344
24	1.43	.43	-.58	2.34	.614	.661
25	1.67	.20	-1.09	1.88	.593	.568
26	1.43	0	-1.42	1.90	.674	.557
27	1.47	.01	-1.22	2.20	.702	.498
28	1.12	.17	-.85	3.40	.723	.617
29	1.31	.17	-.88	2.67	.671	.620
30	1.58	.19	-1.07	1.90	.562	.657
31	1.44	.28	-.73	1.85	.527	.260
32	1.36	.20	-.92	2.71	.515	.252
33	1.16	-.15	-1.01	2.54	.596	.268
34	.86	-.27	-1.20	2.68	.544	.353
35	1.41	-.18	-1.42	1.59	.471	.351
36	1.94	-1.52	-3.45	.71	.401	.048
37	2.17	-.01	-2.00	.89	.451	.532

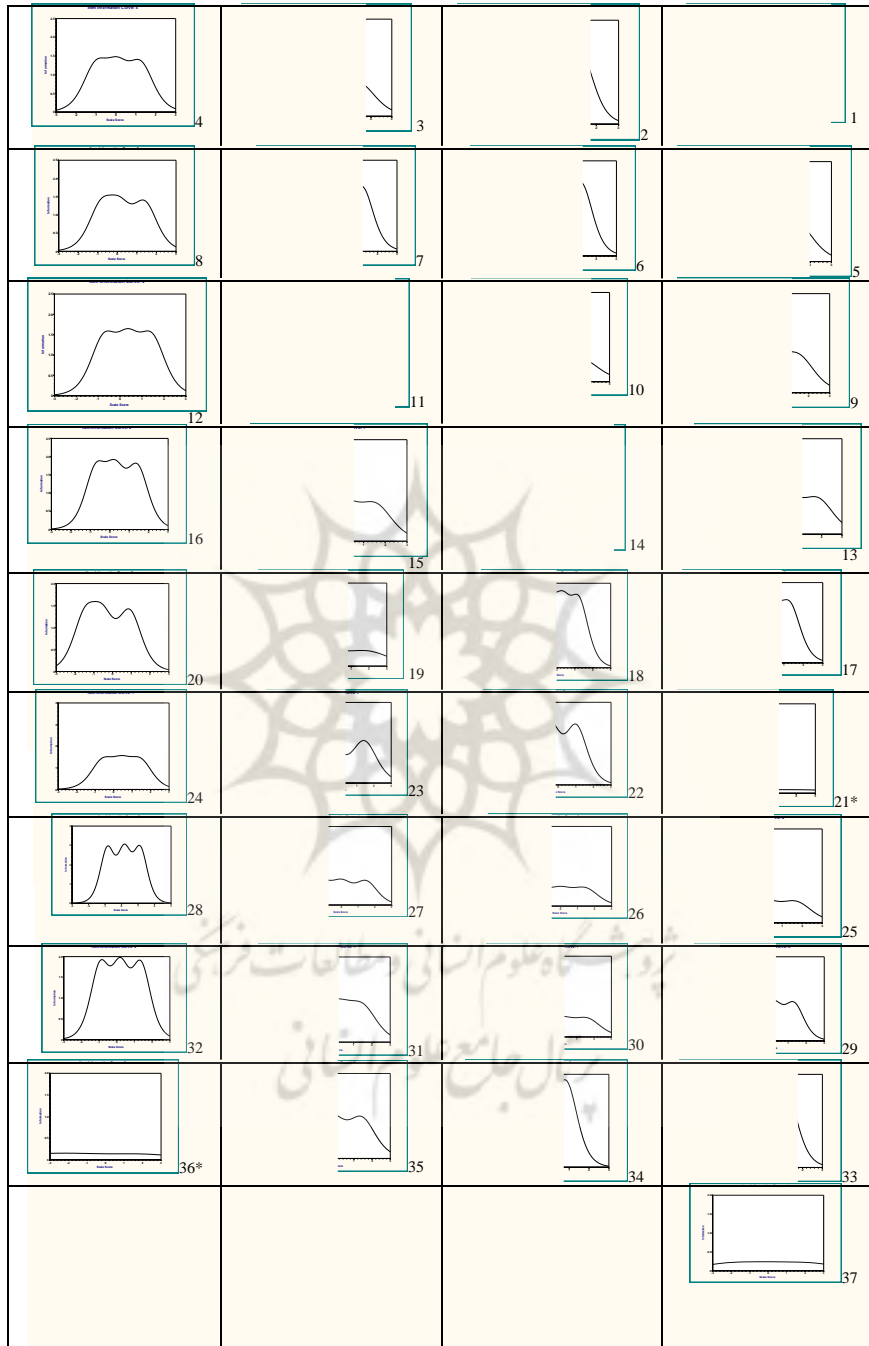


Figure 2. Information functions of items

شکل ۲. توابع آگاهی گویه‌های ابزار

۰/۳ تا ۰/۶ قرار دارد، این مقدار نشان می‌دهد که سؤال‌ها در سطح متوسط قرار دارند. ضرایب تمیز سؤال‌ها در نظریه کلاسیک بین ۰/۳ تا ۰/۵ است. در نظریه سؤال پاسخ ضرایب دشواری و تمیز سؤال‌ها تا حدی متفاوت از نظریه کلاسیک به دست آمده است. همان‌گونه که از نتایج جدول ۸ پیداست دامنه پارامتر تمیز سؤال از ۰/۵۵ تا ۳/۴۰ در نوسان است؛ سؤالانی که پارامتر تمیز کمتر از ۰/۸۰ دارند، سؤال‌ات مناسبی محسوب نمی‌شوند؛ پارامتر تشخیص دو سؤال در نظریه سؤال پاسخ کمتر از ۰/۸ است (سؤال شماره ۲۱ و سؤال ۳۶). سه شاخص بعدی مربوط به آستانه‌های گزینه‌ها است. تابع آگاهی آزمون و به تبع آن منحنی آگاهی آزمون نقش و سهم آزمون را در تعیین سطوح صفت آزمودنی‌ها نشان می‌دهد و در واقع با استفاده از این تابع و منحنی ناشی از آن می‌توان از میزان آگاهی‌دهندگی آزمون از آزمودنی‌ها اطلاع کسب کرد (Rabaniparsa, Mesrabadi & Yarmohammadzade, 2019).

بررسی منحنی ویژگی گویه‌ها نشان داد که در تمامی گویه‌ها پارامترهای آستانه طبقه سوم (خیلی زیاد) به خوبی برآورد شده‌اند. این نتیجه حکایت از آن دارد که بیشترین گزینه‌ای که توسط آزمودنی‌ها انتخاب شده است، گزینه خیلی زیاد است؛ در تمامی گویه‌ها پارامترهای آستانه طبقه دوم (زیاد) در حد متوسط برآورد شده و این بدان معناست که تعداد کمتری از افراد این گزینه را انتخاب کرده‌اند و در انتها در تمامی گویه‌ها پارامترهای آستانه اول (کم) برآورد نسبتاً مطلوبی ندارند و کمتر از دو گزینه دیگر انتخاب شده‌اند. در نهایت هر اندازه که پارامترهای آستانه یک گویه بیشتر و به سمت مثبت باشد بیانگر این موضوع است که این گویه‌ها با صفت مکنون رابطه بالاتری دارند. در جدول ۹ نمرات استاندارد برای هریک از نمرات خام محاسبه شده است.

جدول ۹. نمره‌های هنجار پرسشنامه تشخیص ارزشیابی کاذب

Table 9: Normal scores of Pseudo Evaluation Questionnaire

نمره خام	نمره Z	نمره T	نمره خام	نمره Z	نمره T
35	-3.11	18.89	88	-0.17	48.33
36	-3.06	19.44	89	-0.11	48.89
37	-3.00	20.00	90	-0.06	49.44
38	-2.94	20.56	91	0.00	50.00
39	-2.89	21.11	92	0.06	50.56
40	-2.83	21.67	93	0.11	51.11
41	-2.78	22.22	94	0.17	51.67
42	-2.72	22.78	95	0.22	52.22

52.78	0.28	96	23.33	-2.67	43
53.33	0.33	97	23.89	-2.61	44
53.89	0.39	98	24.44	-2.56	45
54.44	0.44	99	25.00	-2.50	46
55.00	0.50	100	25.56	-2.44	47
55.56	0.56	101	26.11	-2.39	48
56.11	0.61	102	26.67	-2.33	49
56.67	0.67	103	27.22	-2.28	50
57.22	0.72	104	27.78	-2.22	51
57.78	0.78	105	28.33	-2.17	52
58.33	0.83	106	28.89	-2.11	53
58.89	0.89	107	29.44	-2.06	54
59.44	0.94	108	30.00	-2.00	55
60.00	1.00	109	30.56	-1.94	56
60.56	1.06	110	31.11	-1.89	57
61.11	1.11	111	31.67	-1.83	58
61.67	1.17	112	32.22	-1.78	59
62.22	1.22	113	32.78	-1.72	60
62.78	1.28	114	33.33	-1.67	61
63.33	1.33	115	33.89	-1.61	62
63.89	1.39	116	34.44	-1.56	63
64.44	1.44	117	35.00	-1.50	64
65.00	1.50	118	35.56	-1.44	65
65.56	1.56	119	36.11	-1.39	66
66.11	1.61	120	36.67	-1.33	67
66.67	1.67	121	37.22	-1.28	68
67.22	1.72	122	37.78	-1.22	69
67.78	1.78	123	38.33	-1.17	70
68.33	1.83	124	38.89	-1.11	71
68.89	1.89	125	39.44	-1.06	72
69.44	1.94	126	40.00	-1.00	73
70.00	2.00	127	40.56	-0.94	74
70.56	2.06	128	41.11	-0.89	75
71.11	2.11	129	41.67	-0.83	76
71.67	2.17	130	42.22	-0.78	77
72.22	2.22	131	42.78	-0.72	78
72.78	2.28	132	43.33	-0.67	79
73.33	2.33	133	43.89	-0.61	80
73.89	2.39	134	44.44	-0.56	81
74.44	2.44	135	45.00	-0.50	82
75.00	2.50	136	45.56	-0.44	83
75.56	2.56	137	46.11	-0.39	84
76.11	2.61	138	46.67	-0.33	85
76.67	2.67	139	47.22	-0.28	86
77.22	2.72	140	47.78	-0.22	87

افرادی که نمره آن‌ها در بازه ۶۰-۳۵ باشد، «در محدوده پایین ارزشیابی کاذب» قرار می‌گیرند؛ افرادی که نمره آن‌ها در بازه ۸۷-۶۱ باشد، «در معرض ارزشیابی کاذب» قرار

می‌گیرند؛ افرادی که نمره آن‌ها در بازه ۸۸-۱۱۴ باشد، «در محدوده نسبتاً بالای ارزشیابی کاذب» طبقه‌بندی می‌شوند و افرادی که نمره کسب‌شده آن‌ها در بازه ۱۱۵-۱۴۰ باشد، «در محدوده بحرانی ارزشیابی کاذب» قرار می‌گیرند. با توجه به جدول شماره ۹ و هنجارهای محاسبه شده، در صورتی که مجموع امتیازات کسب شده از پرسشنامه ارزشیابی کاذب که از طریق خودگزارشی معلم تکمیل می‌شود، معادل نمره خام ۶۰ باشد، با توجه به چارک‌های گزارش‌های شده، او در مقایسه با ۲۵ درصد از افرادی که در محدوده پایین ارزشیابی کاذب قرار گرفته و کمتر تحت تاثیر عوامل موثر در بروز ارزشیابی کاذب قرار دارند، در وضعیت پرخطرتری قرار دارد؛ به عبارت دیگر این معلم به لحاظ تاثیرپذیری از عوامل موثر در بروز ارزشیابی کاذب از ۷۵ درصد گروه معلمان، وضعیت بهتری دارد.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف ایجاد ابزار اندازه‌گیری برای سنجش کیفیت ارزشیابی‌ها در مدارس ابتدایی از نظر رواج فرهنگ ارزشیابی کاذب انجام شد. منظور از ارزشیابی کاذب فرآیندی است که به جای ارزشیابی معتبر در مدارس روی می‌دهد؛ به عبارت دیگر هر آنچه که باعث شود نتایج ارزشیابی بیش‌برآورد و یا کم‌برآورد شده و نتایج از آنچه که حقیقت داشته فاصله بگیرد مصداق ارزشیابی کاذب است. بدین‌منظور از روش تحقیق آمیخته با طرح اکتشافی استفاده گردید؛ به این شکل که ابتدا در بخش کیفی پژوهش، فن مصاحبه نیمه-ساختارمند، مشاهده، یادداشت‌برداری و بررسی اسناد به‌کار گرفته شد که ماحصل آن مورد تحلیل قرار گرفت و نتایج آن مبنایی برای تنظیم گویه‌های پرسشنامه قرار گرفت. در مطالعه دوم از روش تحقیق کمی به منظور بررسی ویژگی‌های روان‌سنجی ابزار استفاده شد. بدین شکل که بر اساس نتایج بخش کیفی، ابزار اندازه‌گیری در قالب یک پرسشنامه تدوین شد؛ در اجرای مقدماتی با استفاده از ضریب پایایی و بررسی روایی اولیه و اعمال اصلاحات پیشنهادی مجوز اجرای پرسشنامه در فضای نمونه صادر گردید. نتایج حاصل از بخش کیفی پژوهش نشان‌دهنده آن است که مؤلفه‌هایی که به طور بالقوه قابلیت تحریف نتایج ارزشیابی را دارند شامل چهار مقوله کلی است که هر یک از مقوله‌ها از ترکیب چندین مفهوم احصا شده است. در بخش کمی نیز با انجام تحلیل عاملی در مرتبه اول برخی از مفاهیم (خرده مؤلفه‌ها) حذف شد ولی در تعداد مقوله‌های (مؤلفه‌ها) تغییری رخ نداد و بر این اساس ابزار

نهایی کیفیت ارزشیابی در مدارس ابتدایی اصلاح شد؛ شاخص‌های برازش ابزار تدوین شده نیز نشان می‌دهند که مقدار نسبت مجذور خی به درجه آزادی برابر با $2/133$ و شاخص ریشه دوم میانگین مربعات باقیمانده برابر با $0/029$ است و مابقی شاخص‌ها بالای ۹. برآورد شده‌اند که این نشان از برازش مطلوب دارد.

در ادامه نیز به بررسی توابع آگاهی گویه‌ها و کل ابزار پرداخته شد؛ توابع آگاهی بیشینه دقت برآورد سؤال را به ازای بازه توانایی ۳- تا ۳+ نشان می‌دهند. با توجه به توابع آگاهی گویه‌های ابزار می‌توان سؤالات پرسشنامه را در سه گروه پرسش‌هایی که آگاهی بالایی دارند، پرسش‌هایی که آگاهی متوسطی دارند و پرسش‌هایی که آگاهی بالایی دارند دسته‌بندی کرد. پرسش‌هایی که آگاهی بالایی دارند، پرسش‌هایی هستند که در قسمت‌هایی از طیف توانایی دقت برآورد مطلوبی دارند و ضرایب تمیز این سؤالات در حد مطلوبی برآورد شده است. پرسش‌های ۲، ۴، ۶، ۷، ۸، ۱۱، ۱۲، ۱۴، ۱۶، ۱۷، ۱۸، ۲۸، ۳۲ و پرسش ۳۴ آگاهی بالایی دارند. پرسش‌هایی که آگاهی متوسطی دارند، پرسش‌هایی هستند که نسبت به پرسش‌های قبلی دقت برآورد پایین‌تری دارند. اما میزان آگاهی آن‌ها در حد قابل قبول است. سؤالات ۱، ۳، ۵، ۹، ۱۰، ۱۳، ۱۵، ۲۰، ۲۲ و ۲۳؛ ۲۴، ۲۵، ۲۶، ۲۷، ۲۹، ۳۰، ۳۱، ۳۳ و سؤال ۳۵ از میزان آگاهی متوسطی برخوردارند. پرسش‌هایی که آگاهی پایینی دارند پرسش‌هایی هستند که منحی آگاهی آن‌ها تقریباً موازی با محور افقی دستگاه مختصات (یعنی برای بازه‌های مختلف توانایی، آگاهی تقریباً برابر دارند) و مقدار عددی آن‌ها پایین است. از طرفی عموماً این سؤالات ضریب تمیز پایینی نیز دارند و در این ابزار سؤالاتی که آگاهی پایینی دارند سؤالات ۱۹ و ۲۱ و ۳۶ و ۳۷ هستند. از مجموع توابع آگاهی گویه‌ها می‌توان استنباط کرد بیشتر گویه‌ها دقیق‌ترین برآورد و بیشینه آگاهی را در بازه‌های توانایی میانی (۱- تا ۱+) به خود اختصاص داده‌اند. در نهایت تابع آگاهی کل ابزار برآورد شده است که سطح کلی تابع آگاهی ابزار ساخته شده بالاتر از تابع آگاهی یک سؤال به تنهایی است؛ ازین رو یک آزمون نسبت به یک سؤال منفرد توانایی را با دقت بیشتری اندازه می‌گیرد و هرچه تعداد سؤال‌های آزمون بیشتر باشد، مقدار آگاهی آن نیز بزرگتر است؛ بر اساس نمودار تابع آگاهی کل، نظرات افراد با طیف تتای ۱- تا ۱+ با دقت بیشتری برآورد می‌شوند. در مجموع نتایج این پژوهش نشان‌دهنده آن است که پرسشنامه تدوین‌شده دارای ویژگی‌های روان‌سنجی مناسبی برای اندازه‌گیری میزان برخورداری مدارس از ارزشیابی کاذب است. از این رو برای تشخیص

اینکه تا چه میزان فرهنگ ارزشیابی کاذب در مدارس ابتدایی رواج یافته‌است می‌توان از این ابزار استفاده کرد و اقدامات لازم را به منظور جلوگیری و رفع موانع به‌عمل آورد.

نسخه خودگزارشی معلمان: پرسشنامه میزان‌شده برای تشخیص ارزشیابی کاذب در مدارس ابتدایی

ردیف	چه اندازه بر ارزشیابی شما (معلم) از عملکرد تحصیلی دانش‌آموزان تاثیر می‌گذارد؟	هر یک از گویه‌های زیر تا	خیلی زیاد	زیاد	کم	خیلی کم
۱	دخالت دادن توصیه‌های مدیر و سایر عوامل مدرسه در نتایج ارزشیابی دانش‌آموزان					
۲	افت تمرکز به دلیل حجم بالای فعالیت‌ها					
۳	شناخت قبلی نسبت به شرایط خانوادگی دانش‌آموز					
۴	ناکافی بودن ملاک‌های ارزشیابی برای جمع‌بندی عملکرد دانش‌آموزان					
۵	اثرگذاری نظر معلمان دیگر نسبت به دانش‌آموزان در ارزشیابی					
۶	تاثیرپذیری از درخواست‌ها و فشارهای والدین					
۷	نامناسب بودن چیدمان برنامه درسی					
۸	افت انگیزه ناشی از حجم زیاد فعالیت‌ها					
۹	آشنایی ناکافی با شیوه‌ها و ابزارهای ارزشیابی					
۱۰	تناسب نداشتن محتوای درسی با امکانات برای تدریس					
۱۱	ابهام در مفاد آیین‌نامه‌ها و مقررات ابلاغی مربوط به ارزشیابی دانش‌آموزان					
۱۲	نداشتن راهنمای دقیق برای اجرای ارزشیابی دانش‌آموزان					
۱۳	اثرگذاری شناخت قبلی از دانش‌آموز بر قضاوت درباره عملکرد او					
۱۴	اثر یکنواختی و اشیاع‌زدگی فعالیت‌های شغلی					
۱۵	اثر حجیم بودن فعالیت‌های شغلی					
۱۶	اثرات ناشی از توجه نامتوازن به دروس					
۱۷	اثر ناشی از کیفیت پایین آموزش‌ها در دوره‌های ضمن خدمت					
۱۸	اثر ناشی از ضعف عملکرد معلمان پایه‌های قبلی در آموزش پیش‌نیازها					
۱۹	کمبود یا عدم دسترسی به امکانات، فضا و ابزار مناسب برای ارزشیابی					
۲۰	میهم و کلی بودن ملاک‌ها برای جمع‌بندی عملکرد دانش‌آموزان					
۲۱	عینیت نداشتن استانداردها برای جمع‌بندی عملکرد دانش‌آموزان					
۲۲	کم توجهی به رعایت اصول ارزشیابی در فرایند اجرا					
۲۳	اثر ناآشنا بودن جامعه فرهنگی با تحولات آموزش و تاثیر آن بر کیفیت ارزشیابی					
۲۴	اثر ناآشنا بودن خانواده‌ها با تحولات حوزه ارزشیابی و تاثیر آن بر کیفیت قضاوت‌های معلم					
۲۵	اثر وضعیت استخدامی روی نتایج ارزشیابی از دانش‌آموزان					
۲۶	کم توجهی به برقراری تناسب بین تکالیف درسی با نیاز دانش‌آموزان					
۲۷	اثر زمان و حجم کلاس‌ها در ارزشیابی					
۲۸	سطح تناسب تعداد و نسبت دانش‌آموزان به معلم					
۲۹	اثرگذاری پیشینه فرهنگی دانش‌آموز بر قضاوت درباره عملکرد او					
۳۰	سطح دسترسی شما به امکانات، ابزارها و منابع اطلاعاتی و آموزشی					
۳۱	اثرگذاری رقابت با معلمان هم‌پایه در مدرسه					
۳۲	ترس از طرد شدن به دلیل ناهمسویی با سیاست‌های مدیر مدرسه و تاثیر آن بر کیفیت ارزشیابی دانش‌آموزان					
۳۳	نگاه مچ‌گیرانه به ارزشیابی و کارکردهای آن					
۳۴	ناآگاهی از سطوح هدف‌های آموزشی هر درس (عاطفی، شناختی و روانی حرکتی) و تاثیر آن بر کیفیت ارزشیابی					
۳۵	آشنایی ناکافی با سبک‌های یادگیری دانش‌آموزان و تاثیر آن بر کیفیت ارزشیابی					

سهم مشارکت نویسندگان: در پژوهش حاضر نویسنده دوم، به عنوان استاد راهنما، نظارت و راهبردی روند کلی پژوهش و تدوین و نهایی‌سازی اصلاحات مقاله را بر عهده داشته‌اند. نویسنده اول در تدوین طرح تحقیق، فرآیند گردآوری، تحلیل و تفسیر یافته‌ها و نگارش متن مقاله را بر عهده داشته و در مجموع نتیجه‌گیری از یافته‌ها و بسط و تفسیر به صورت مشترک و با بحث و تبادل نظر کلیه همکاران و با همراهی نویسنده سوم، به عنوان استاد مشاور پایان‌نامه، انجام شد.

تضاد منافع: نویسندگان اذعان دارند که در این مقاله هیچگونه تعارض منافی وجود ندارد.

منابع مالی: پژوهش حاضر از هیچ موسسه و نهادی حمایت مالی دریافت نکرده و کلیه هزینه‌ها در طول فرآیند اجرای پژوهش بر عهده پژوهشگران بوده است.

تشکر و قدردانی: پژوهش حاضر بدون همکاری مشارکت کنندگان امکان‌پذیر نبود؛ بدینوسیله از کلیه مشارکت کنندگان تقدیر و تشکر به عمل می‌آید.

References

- Abedi, A. (2011). Challenges and Strategies for Effective Evaluation of Social Programs, *Journal of Research*, 3(2), 139-155. [Persian]
- Bazargan, A. (2012). *Educational Evaluation*, Samt, Tehran. [Persian]
- Desautels, G., & Jacob, S. (2012). The ethical sensitivity of evaluators: Aqualitativestudy using a vignette design. *Evaluation*, 18(4), 437-450.
- Dewhirst, C. (2019). *Evaluation of learning and teaching*. Queens University Belfast. Northern Ireland.
- Ghaemipour, M., Karami, M., & Khandaghi, M. (2015). Dissemination of Evaluation Standards in Educational Evaluation Research Report: *Exploring Master's Theses*; 5(9), 101-128. [Persian]
- Ghorbankhani, M., & Salehi, K., (2017). Representation the characteristics of the successful professors in the virtual education in Iran's higher education system from the viewpoint of students and professors: a phenomenological study. *Journal of Technology of Education*, 11(4), 327-347. [Persian]
- Ghorbankhani, M., & Salehi, K. (2017). Representation Challenges of Virtual Training in Iran's Higher Education System: A Study of Phenomenological Approach. *Information and Communication Technology in Educational Sciences*, 7(2), 26, 123-148. [Persian]
- Ghorbankhani, M., Salehi, K., & Moghadamzadeh, A. (2019). Identifying effective factors in shaping the culture of Psude evaluation in primary schools. *Educational Measurement and Evaluation Studies*, 9(26), 29-69. [Persian]
- Habibi, H., Pardakhtchi, M., Abolghasemi, M., & Ghahremani, M. (2013). Investigating the school-centered learning atmosphere and its relationship with school effectiveness. *Teaching and Learning Researches*, 20(3), 101-118. [Persian]
- Khosrojerdi, N., & Mahmoudi, M. (2014). School is a safe place to live and learn, National Conference on Architecture, *Urban Planning and*

Sustainable Development with a focus on reading Iranian identity. <https://civilica.com/doc/346932> [Persian]

- Lai Fong Cheng, A., & Keung Yau, H. (2011). Principals' and teachers' perceptions of quality management in Hong Kong primary schools. *Quality Assurance in Education*, 19(2), 170-186.
- Lee, C., Oh, P. S., Ang, A., & Lee, G. (2014). Holistic assessment implementation in Singapore primary schools—part I: Using assessment to support the learning and development of students. In *International Association of Educational Assessment Conference, Singapore*.
- Mahdiyoun, R., Pahang, N., & Yarigholi, B. (2017). Identifying the factors affecting the quality of schools and examining its current status: Mixed Method, *School Management*, 5(1), 173-193. [Persian]
- Mohammadzadeh, Z., & Salehi, K. (2017). The process of evaluating academic achievement from the perspective of the Holy Quran. *Journal of Research in Islamic Education*. 25(37), 100-79. [Persian]
- Nxumalo, Z. F. (2007). The role of continuous assessment in primary school. Thesis of Master of education (*Doctoral dissertation*). University of Zululand, Faculty of Education, Department of Foundations of Education.
- Porzahir, A. (2016). *Introduction to Educational Planning*, Agah, Tehran. [Persian]
- Rabbani Parsa, E., Mesrabadi, J., & Yarmohammadzade, P. (2019). Build and Standardization of Critical Thinking, *Educational Measurement*, 9(34), 91-114. [Persian]
- Salehi, K., & Golafshani, N. (2010). Commentary: Using Mixed Methods in Research Studies—an Opportunity with Its Challenges. *International Journal of Multiple Research Approaches*, 4(3), 186-191. <https://doi.org/10.5172/mra.2010.4.3.186> [Persian]
- Salehi, K., Bazargan, A., Sadeghi, N., & Shokoohi-Yekta, M. (2015). Representing the Teacher's Perceptions and Lived Experiences of the Possible Damages for Implementing the Descriptive Evaluation Scheme in Primary Schools, *Journal of Measuring and Educational Evaluation Studies*, 5(9), 59-99. [Persian]
- Samuel, K. H. (2008). Continuous assessment and lower attaining pupils in primary and junior secondary schools in Ghana (*Doctoral dissertation*). University of Birmingham.
- Sarmad, Gh. (2010). *Teaching methods and teacher art*. Avaye noor, Tehran. [Persian]
- Seif, A. (2012). *Educational measurement, assessment and evaluation*. Doran, Tehran. [Persian]
- Shabani, H. (2017). *Educational skills*, Samt. Tehran. [Persian]
- Shahdosti, L. (2011). *The role of variables in home environment, student background, school atmosphere and teaching process in the mathematical development of third grade middle school students based*

- on clean studies (2007). Master Thesis*, Shahid Madani University of Azerbaijan, Tabriz, Iran. [Persian]
- Shoarinejad, A. (2015). *Psychology of Education and Teaching*, Second Edition, Etelaat. Tehran. [Persian]
- Sonpal-Valias, N. (2009). *MODULE 1 Outcome Evaluation: Definition and Overview*, the Vocational and Rehabilitation Research Institute.
- Stufflebeam, D. L. (2001). The metaevaluation imperative. *The American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L., & Shinkfield, A. (2012). *Evaluation, Theory, Models and Application*. Jossey-Bass Willy.
- Tavakoli, A., Ghanbaripanah, A., Ghorbankhani, M., & Zivari, M. (2019). Psychometric properties of the scale of study habits in female students. *Psychometry*, 8(29), 83-105. [Persian]
- Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, 8 (1), 60-80.

