Research Paper

The Position of Belief Functions Structure in Performance Evaluation Measures: a New Approach

Hosein Nahid [1] , Davoud khani [*2]

[1] Department of industrial engineering, Payame Noor University, Tehran, Iran.
[2] Department of Management, Payam Noor University, Tehran, Iran.

10.22080/SHRM.2023.4393

Received:
June 22, 2023
Accepted:
September 6, 2023
Available online:
October 5, 2023

Keywords:
Performance evaluation, Evaluation scale, Belief functions, Evidence Theory

Abstract

Performance evaluation is one of the management processes in the organization, the results of which have a great impact on human resources policies. In the performance evaluation process, the nature of the evaluated phenomenon and the evaluation process are factors that create various types of uncertainty in the evaluation results and affect the accuracy and validity of the results. Since the common performance evaluation scales do not have the appropriate capability of modeling and dealing with uncertainty, in many cases the results obtained from these models have been evaluated as weak and invalid. Among the methods of dealing with uncertainty, the theory of evidence or the theory of belief functions is a suitable structure for gathering inconclusive opinions from evaluators and it can simultaneously face the uncertainty caused by the variability and uncertainty of the lack of knowledge. In this paper, a new way to get opinions from evaluators based on belief functions structure was presented and the advantages and improvements of the presented model were compared to the common metrics used in the performance evaluation process. The results of the implementation of the model in evaluating the performance of university lecturers showed that when simple and clear questions are used in evaluations, there is no significant difference between the results obtained from different evaluation methods. The advantages of the belief functions structure in the face of uncertainty can facilitate the possibility of expressing opinions for evaluators and improve the final results of the evaluation.

Extended abstract

1. Introduction

In common performance evaluation models, Likert Scales, Fuzzy Linguistic Scales (FLS), and Visual Analogue Scales (VAS) are used, and evaluators are asked to express their opinion by choosing an option on the evaluation scale.

This type of data is associated with uncertainty due to the existence of several factors such as the existence of fuzzy concepts (the complexities of human behavior), contradictory evidence (different human behavior in different situations), and evaluators' lack of enough information. For example in the performance evaluation process, there are situations where the evaluators do not have enough information in all evaluation criteria to express their opinion with full confidence. Therefore, this limited

[*]**Corresponding Author:** Davoud khani
Address: Department of Management, Payam Noor University, Tehran, Iran
Email: davoud_khani@pnu.ac.ir

information cannot be accurately obtained using conventional evaluation scales.

Although the use of these types of scales is very common in performance evaluation models, the uncertainty in the data obtained from these scales has caused numerous errors in the performance evaluation process.

In this regard, several studies have been conducted to investigate various dimensions of error and uncertainty in the performance evaluation process.

The uncertainty in the evaluation data requires an appropriate tool for dealing with uncertainty to be included in the performance evaluation models. One of the effective tools in this field is Evidence Theory.

The Evidence Theory, which is also referred to as Dempster–Shafer theory (DST) or theory of Belief Functions, is a general framework for dealing with uncertainty and inference based on uncertain information. In this theory, the belief structure is a powerful and flexible tool for modeling the probabilistic uncertainty (Aleatory uncertainty) and also the uncertainty due to knowledge deficiency (epistemic uncertainty). In evidence theory, it is possible to combine shreds of evidence obtained from different sources using evidence combination rules, in which the conflict between pieces of evidence can also be addressed. This theory has some desirable properties and also has been widely used in many domains such as risk analysis, sensor data fusion, diagnosis of faults or diseases, and group decision-making.

In this paper, by looking at this problem from the Evidence Theory point of view, Belief Functions Structure as a flexible and powerful framework in modeling different types of uncertainty (uncertainty caused by changeability and uncertainty caused by lack of knowledge) is proposed as a solution to deal with uncertainty in the performance evaluation process.

2. Research Methods

Two questionnaires were used to evaluate the performance of a university lecturer. Each questionnaire was implemented in four modes using different evaluation scales including Likert Scales, Fuzzy Linguistic Scale (FLS), Visual Analogue Scale (VAS), and Belief Functions Structure. The first questionnaire contained five simple and clear questions and the second questionnaire contained three complex questions. In the next step, after summarizing the results of each questionnaire, a survey was conducted regarding the acceptability of the obtained results.

For analyzing the collected data, a one-way ANOVA parametric test and Friedman's non-parametric test were used.

3. Results

The results showed that in the first questionnaire (simple questions), there is no significant difference between the results obtained from the four evaluation scales. However, this difference is significant in the second questionnaire (complex questions) in which there was more ambiguity and uncertainty. According to the significant difference identified, in the next step, the acceptability of the results was checked and it was found that the results obtained by the Belief Functions Structure have higher acceptability than other evaluation scales.

In addition, in terms of the ease of working with evaluation scales, the Likert Scale in simple questions and the Belief Functions Structure in complex questions were ranked high.

4. Conclusion

The advantages of Belief Functions Structure in the face of uncertainty can facilitate the declaration of opinion for evaluators and improve the final results of the evaluation process. Therefore, in evaluations and surveys where there is uncertainty in the data due to the nature of the questions, an appropriate tool should be used to deal with the uncertainty in the assessment data. The findings of this research show that the proposed evaluation model based on Belief functions can be a good choice in this situation.

Funding:
There is no funding support.

Authors' contribution:
All authors are equally involved in the preparation and writing of the article.

Conflict of interest:

The authors declared no conflict of interest.

Acknowledgments:

The authors appreciate all the scientific consultants in this article as well as the employees of the General Directorate of Information Technology of the Tax Affairs Organization for their cooperation.

علمی

جایگاه ساختار توابع باور در سنجه‌های ارزیابی عملکرد: رهیافتی نوین

حسین ناهید[۱] داوود خانی[۲*]

[۱]. دانشکده فنی و مهندسی، گروه مهندسی صنایع، دانشگاه پیام نور، تهران، ایران.
[۲]. گروه مدیریت، دانشگاه پیام نور، تهران، ایران

10.22080/SHRM.2023.4393

تاریخ دریافت:
۱ تیر ۱۴۰۲
تاریخ پذیرش:
۱۵ شهریور ۱۴۰۲
تاریخ انتشار:
۱۳ مهر ۱۴۰۲

چکیده

در فرایند ارزیابی عملکرد کارکنان، به‌عنوان یکی از فرایندهای مدیریتی مهم حوزه منابع انسانی، «ماهیت پدیده مورد سنجش» و «فرایند ارزیابی»، از عوامل مهم ایجاد عدم قطعیت در نتایج ارزیابی هستند. با توجه به اینکه مقیاس‌های متداول ارزیابی عملکرد از قابلیت مناسب مدل‌سازی و مواجهه با عدم قطعیت برخوردار نیستند، در موارد متعدد نتایج حاصل از این مدل‌ها ضعیف و نامعتبر ارزیابی شده است. در بین روش‌های مواجهه با عدم قطعیت، تئوری شواهد یا تئوری توابع باور، ساختاری مناسب برای گردآوری نظرات غیرقطعی از ارزیاب‌ها بوده و قابلیت مواجهه همزمان با عدم قطعیت ناشی از تغییرپذیری و عدم قطعیت ناشی از نقص دانش را فراهم می‌آورد. لذا در این مقاله شیوه جدیدی مبتنی بر ساختار توابع باور جهت دریافت نظر از ارزیاب‌ها ارائه شده و مزایا و بهبودهای حاصل از ابزار پیشنهادی مورد بررسی قرار گرفته است. برای مقایسه عملکرد ابزار پیشنهادی با سنجه‌های متداولِ مورد استفاده در فرایند ارزیابی عملکرد، دو پرسش‌نامه‌ی محقق ساخته در دو سطح سؤالات ساده و سؤالات پیچیده، جهت ارزیابی عملکرد یک مدرس دانشگاهی تهیه و نظرات ۵ دانشجو در قالب چهار سنجه لیکرت، عبارات زبانی فازی، مشابهت بصری و ساختار توابع باور دریافت شد. جهت بررسی معنادار بودن تفاوت نتایج به‌دست‌آمده، از آزمون پارامتریک «آنالیز واریانس یک‌طرفه با داده‌های تکراری» و جهت بررسی مقبولیت نتایج به‌دست‌آمده، از آزمون‌های ناپارامتریک «فریدمن و ویلکاکسون» استفاده شد. بررسی نتایج به‌دست‌آمده نشان داد زمانی که از سؤالات ساده و شفاف در ارزیابی‌ها استفاده شود تفاوت معناداری بین نتایج به‌دست‌آمده از روش‌های مختلف ارزیابی وجود ندارد اما در پاسخ به سؤالات پیچیده مزیت‌های ساختار توابع باور در مواجهه با انواع عدم قطعیت، می‌تواند امکان اعلام نظر ارزیاب‌ها را تسهیل کرده و نتایج نهایی حاصل از ارزیابی را بهبود دهد.

کلیدواژه‌ها:
ارزیابی عملکرد؛ مقیاس ارزیابی؛
توابع باور؛ تئوری شواهد

۱ مقدمه

ارزیابی عملکرد کارکنان یکی از فرآیندهای مهم حوزه فرآیند مدیریت منابع انسانی در هر سازمان است که با اهداف متنوعی مانند توسعه فرآیند کارکنان، مدیریت حقوق و مزایا، استخدام و جایابی، برنامه‌های ارتقا و جانشینی کارکنان، ارزیابی نیازهای آموزشی، ارزیابی تأثیر دوره‌های آموزشی و یا دریافت بازخوردهایی جهت توسعه و حفظ عملکرد سازمان اجرا می‌شود. نظر به اهمیت موضوع، تاکنون مطالعات فراوانی در حوزه فرآیند ارزیابی عملکرد کارکنان صورت گرفته و سیستم‌ها و مدل‌های گوناگون ارزیابی عملکرد کارکنان ارائه شده است(Turgut & Mert, 2014) . با توجه به اینکه اغلب شاخص‌های ارزیابی عملکرد کارکنان که مرتبط با ویژگي‌هاي فردي و رفتار افراد جنبه فرآیند کیفی دارند و شاخص‌های مناسب کمی برای سنجش آن‌ها وجود ندارد (Golman & Bhatia, 2012)، تقریباً تمامی سیستم‌های متداول ارزیابی عملکرد کارکنان بر اساس نظرسنجی از افراد

* **نویسنده مسئول:** داوود خانی
آدرس: گروه مدیریت، دانشگاه پیام نور، تهران، ایران
ایمیل: davoud_khani@pnu.ac.ir

طراحی شده‌اند و در اکثر آن‌ها از مقیاس رتبه‌بندی[1] جهت ارزیابی عملکرد استفاده می‌شود (Aiken, 1996; Austin & Villanova, 1992). استفاده از سیستم‌های مبتنی بر مقیاس‌های رتبه‌بندی مبتنی بر مقیاس معمول لیکرت تا حدی متداول است که گاهی در تحقیقات حوزه فرآیند مدیریت منابع انسانی و روان‌شناسی سازمانی، ارزیابی عملکرد مترادف با رتبه‌بندی عملکرد[2] در نظر گرفته می‌شود (Brutus, 2010).

در مدل‌های متداول ارزیابی عملکرد، داده‌های ارزیابی به‌طورمعمول با استفاده از مقیاس ارزیابی لیکرت[3]، مقیاس عبارات زبانی فازی[4] و مقیاس شباهت بصری[5] تولید می‌شوند. مقیاس لیکرت که متداول‌ترین نوع مقیاس ارزیابی است و در آن معمولاً از اعداد یا عبارات زبانی متناظر با اعداد جهت رتبه‌بندی و تحلیل پاسخ‌ها استفاده می‌شود. مقیاس عبارات زبانی فازی، کاملاً مشابه مقیاس لیکرت مبتنی بر عبارات زبانی است با این تفاوت که در تحلیل داده‌های حاصل از آن از مفاهیم مرتبط با اعداد فازی متناظر با گزینه‌های ارزیابی استفاده می‌شود. در مقیاس مشابهت بصری طیف پاسخ‌دهی پیوسته بوده و معمولاً بازه یک تا صد را برای اعلام نظر استفاده می‌کنند و ارزیاب می‌تواند هر نقطه‌ای در این طیف را که به نظر او نزدیک‌تر است، انتخاب و اعلام نماید (Lubiano et al., 2016a).

در مدل‌های متداول ارزیابی عملکرد از ارزیاب‌ها خواسته می‌شود تا نظر خود را با مشخص‌کردن یک گزینه بر روی مقیاس ارزیابی اعلام نمایند اما رفتار متفاوت افراد در موقعیت‌های مختلف و پیچیدگی‌های ذاتی رفتارهای انسانی به‌عنوان یکی از مهم‌ترین دلایل بروز عدم قطعیت، باعث شده است تا ارزیاب‌ها نتوانند به‌راحتی و از طریق تعیین یک عدد دقیق بر روی مقیاس ارزیابی، نظر خود را اعلام نماید (Arbaiy & Suradi, 2007; Azwir & Kalinggo, 2019; Golman & Bhatia, 2012; Levy & Williams, 2004; Manoharan et al., 2011). لذا با وجود گستردگی استفاده از مقیاس‌های رتبه‌بندی در مدل‌های ارزیابی عملکرد کارکنان، داده‌های حاصل از این نوع ارزیابی‌ها در معرض انواع خطا و عدم قطعیت است (Golman & Bhatia, 2012; Neogi et al., 2011; Turgut & Mert, 2014). در همین راستا در مطالعات متعدد به تأثیر مقیاس‌های ارزیابی بر روی نظر ارزیاب و طراحی مقیاس ارزیابی مناسب که امکان دریافت نظر صحیح از ارزیاب را تسهیل کند اشاره شده است (Golman & Bhatia, 2012; Hedge & Teachout, 2000; Hoffman et al., 2012; Levy & Williams, 2004; Lubiano et al., 2016a; Manoharan et al., 2011; Ying-Feng & Ling-Show, 2002).

باوجوداین در هیچ‌کدام از روش‌های شناخته‌شده فرآیند ارزیابی عملکرد، مسأله فرآیند عدم قطعیت به‌طور مشخص لحاظ نشده است (Esen et al., 2016; Gürbüz & Albayrak, 2014; Manoharan et al., 2011). اصلی‌ترین رویکرد اتخاذشده در مطالعات مرتبط با افزایش کیفیت داده‌ها در فرآیند ارزیابی عملکرد، که بیشتر با رویکرد روان‌شناسی انجام شده‌اند، ارائه فرآیند اطلاعات تکمیلی به ارزیاب جهت ایجاد درک مشترک و استاندارد نسبت به معیارها و مقیاس ارزیابی در فرآیند ارزیابی عملکرد کارکنان است. ازجمله مهم‌ترین این راهکارها می‌توان به آموزش چارچوب مرجع ارزیابی[6]، مقیاس ارزیابی مبتنی بر نشانه‌های رفتاری[7] و مقیاس ارزیابی چارچوب مرجع[8] اشاره نمود. در راهکار آموزش چارچوب مرجع ارزیابی، به ارزیاب‌ها در مورد معانی معیارهای ارزیابی و مقیاس مورد استفاده در اعلام نظر آموزش داده می‌شود (Hoffman et al., 2012; Woehr & Huffcutt, 1994a). در مقیاس ارزیابی مبتنی بر نشانه‌های رفتاری، مثال‌هایی در مورد مصادیق رفتاری مرتبط با هر کدام از گزینه‌های مشخص‌شده بر روی مقیاس ارزیابی درج می‌شود. در مقیاس ارزیابی چارچوب مرجع، تشریح معیار و مثال‌های مرتبط با آن قبل از معیار ارزیابی ارائه می‌شود (Murphy, 2008a). این راهکارها مزایایی در افزایش کیفیت داده‌های ارزیابی ایجاد می‌نمایند؛ اما پیاده‌سازی آن‌ها با دشواری‌هایی مواجه است. به‌عنوان‌مثال در مقیاس ارزیابی مبتنی بر نشانه‌های رفتاری، اختصاص امتیاز صحیح به هر نشانه فرآیند رفتاری دشوار است و یا راهکار آموزش چارچوب مرجع هزینه‌بر است (Hoffman et al., 2012).

با وجود مزایایی که راهکارهای ذکرشده در بهبود کیفیت داده‌های ارزیابی از طریق ایجاد مبنای ارزیابی استاندارد و شفاف ایجاد می‌نمایند، در این راهکارها به مشکل اساسی ذکرشده در ارتباط با انتخاب یک گزینه فرآیند مشخص از بین چند گزینه بر روی مقیاس ارزیابی و عدم قطعیت ناشی از آن در فرآیند ارزیابی عملکرد کارکنان پرداخته نشده است. رویکرد دیگر جهت افزایش کیفیت داده‌ها در فرآیند ارزیابی عملکرد کارکنان، استفاده از ابزارهای متداول کار با داده‌های غیرقطعی و طراحی مقیاس‌های ارزیابی با قابلیت مواجهه با عدم قطعیت است. راهکارهای ارائه‌شده در این حوزه محدود به کاربرد ابزارها و مفاهیم فازی بوده و در این راستا، تاکنون دو مقیاس ارزیابی شامل مقیاس عبارات زبانی فازی[9] و مقیاس ارزیابی فازی[10] ارائه شده است (Lubiano et al., 2016a).

در مقیاس عبارات زبانی فازی، ارزیاب نظر خود را با مشخص‌کردن یک عبارت زبانی فازی اعلام می‌نماید. در این مقیاس باوجود مدل‌سازی عدم قطعیت با استفاده از عبارات فازی، هنوز دشواری انتخاب یک گزینه مشخص وجود دارد و شک و تردید بین چند گزینه ارزیابی در آن قابل مدل‌سازی نیست. در روش دیگر تحت عنوان مقیاس ارزیابی فازی، ارزیاب می‌تواند نظر خود را با ترسیم یک عدد فازی (معمولاً ذوزنقه‌ای) بر روی بازه ارزیابی مشخص نماید. این روش،

[1] rating scales
[2] performance rating
[3] Likert
[4] Fuzzy Linguistic Scale
[5] visual analogue scale
[6] Frame Of Reference Training (FORT)
[7] Behaviorally Anchored Rating Scales (BARS)
[8] Frame of Reference Scales (FORS)
[9] Fuzzy verbal values
[10] Fuzzy Rating Scale

انعطاف بالایی در نمایش اطلاعات موردنظر ارزیاب و مدل‌سازی عدم قطعیت در فرآیند نظرسنجی دارد؛ اما باوجود مزایای یادشده، این مقیاس ازنظر اجرا و پیاده‌سازی و نیز تحلیل داده‌های حاصل از آن با مشکلاتی مواجه است و به همین دلیل کاربردهای بسیار محدودی داشته است (Lubiano et al., 2016a).

راهکار قدرتمند دیگری که مبتنی بر تئوری شواهد جهت دریافت نظرات و قضاوت‌های توأم با عدم قطعیت توسعه یافته است، انجام ارزیابی در ساختار توابع باور است (Yang & Singh, 1994). ساختار باور نوعی ارزیابی توزیع‌شده با استفاده از سطوح باور است که در ارزیابی یک عامل بر اساس یک معیار مشخص استفاده می‌شود(Ng, 2016). استفاده از این ساختار به افزایش عینیت و قابلیت اعتماد در فرآیند گردآوری داده‌ها و مراحل اولیه فرآیند ارزیابی کمک می‌کند. این ابزار به‌عنوان ابزاری قدرتمند و یکپارچه در مدل‌سازی انواع مختلف عدم قطعیت شامل عدم قطعیت احتمالی و عدم قطعیت ناشی از نقص دانش شناخته شده و در مطالعات متعدد مرتبط با کاربرد تئوری شواهد در تصمیم‌گیری مورد استفاده قرار گرفته است(Xu, 2012). داده‌های حاصل از ساختار باور در قالب توابع جرم باور بوده و تحلیل آن‌ها بر اساس مفاهیم تئوری شواهد انجام می‌شود.

باوجود قابلیت‌های فراوان، تئوری شواهد و ساختار باور در دریافت نظرات و قضاوت‌های توأم با عدم قطعیت، تاکنون در حوزه فرآیند ارزیابی عملکرد کارکنان مورد استفاده قرار نگرفته است. دانشگاه به‌عنوان رکن تولید علم و تغییر در هر کشور، محل تبلور بسیاری از دغدغه‌های حاکمیتی در حوزه فرآیند توسعه و تولید سرمایه فرآیند فکری است؛ اصلاح نظام تولید علم بدون توجه به سازوکارها، فلسفه و ارکان اصلی آن و نیز ارزش‌یابی عملکرد آموزشی، پژوهشی و خدماتی مراکز و مؤسسات آموزش عالی مقدور نخواهد بود (Cotton et al., 2017). ازاین‌رو در این مقاله با رجوع به یک نمونهٔ دانشگاهی، این دیدگاه که استفاده از ساختار باور در مرحلهٔ دریافت نظر از ارزیاب‌ها و در مرحلهٔ بعد تجمیع داده‌های حاصل از آن با استفاده از قوانین ترکیب شواهد، می‌تواند به تسهیل فرآیند نظرسنجی و کسب نتایج دقیق‌تر و با اعتبار بیشتر نسبت به روش‌های معمول ارزیابی عملکرد منتهی گردد به بوته آزمون گزارده خواهد شد. هدف مقالهٔ حاضر بررسی قابلیت‌های ساختار باور و تئوری شواهد در افزایش دقت نتایج در فرآیند ارزیابی عملکرد کارکنان و تسهیل اعلام نظر ارزیاب‌ها است. به‌عبارت‌دیگر آیا ساختار توابع باور می‌تواند به‌عنوان ابزاری با اثربخشی بیشتر در قیاس با شیوه‌های متداول ارزیابی عملکرد و سنجش نظرات افراد در نظر گرفته شود؟

لذا در ادامه، بررسی پیشینهٔ تحقیقات مرتبط با موضوع پژوهش انجام می‌گیرد. در بخش سوم روش انجام پژوهش و نحوهٔ مقایسه عملکرد مدل ارائه‌شده با مدل‌های متداول ارزیابی عملکرد تشریح می‌گردد. در بخش چهارم نتایج حاصل از اجرای میدانی مدل ارائه شده و مورد بررسی قرار می‌گیرد و درنهایت در بخش پنجم جمع‌بندی و نتایج حاصل از انجام این پژوهش ارائه می‌گردد.

۲ مبانی نظری و پیشینهٔ پژوهش

۲,۱ مبانی تئوری توابع باور (تئوری شواهد)

۲,۱,۱ معرفی و کاربردها

تئوری شواهد که از آن به تئوری دمستر - شیفر (دمستر ۱۹۶۷ و شیفر ۱۹۷۶) و یا تئوری توابع باور نیز یاد می‌شود به‌عنوان ابزاری انعطاف‌پذیر که قابلیت مواجهه با انواع عدم قطعیت را دارد و در زمینه‌های مختلف مورد استفاده قرار گرفته است در این حوزه مطرح است. این تئوری در ابتدا جهت دریافت نظرات توأم با عدم قطعیت از افراد ایجاد شد و ساختار دریافت نظرات در این تئوری به نحوی است که امکان اعلام نظر بسیار منعطف و راحت را برای ارزیاب‌ها فراهم می‌آورد اما به جهت انعطاف‌پذیری و قابلیت مدل‌سازی انواع عدم قطعیت در زمینه‌های مختلف مهندسی، پزشکی و تصمیم‌گیری مورد استفاده قرار گرفته است (Ding et al., 2012; Huang et al., 2021; Maseleno et al., 2015; Oberkampf et al., 2002; Qiang & Deng, 2022b).

تئوری شواهد به‌عنوان ابزاری که قابلیت دریافت انواع الگوهای شواهد را داشته و توانایی مواجهه همزمان با عدم قطعیت تصادفی و عدم قطعیت شناختی را دارد، ابزاری مناسب جهت استفاده در حوزهٔ ارزیابی عملکرد است. ساختار توابع باور در این تئوری این امکان را فراهم می‌آورد تا ارزیاب‌ها به‌راحتی نظرات و قضاوت‌های توأم با عدم قطعیت خود را بیان نمایند. ساختار باور نوعی ارزیابی توزیع شده با استفاده از سطوح باور است که برای ارزیابی عملکرد یک عامل بر اساس یک معیار استفاده می‌شود (Ng, 2016). این مفهوم در مطالعات متعدد در حوزه‌های تصمیم‌گیری مورد استفاده قرار گرفته است (Fu et al., 2015; Jin et al., 2015). در این روش اعلام نظر، ارزیاب قادر است به‌جای تجمیع شواهد و اطلاعات متناقض در ذهن خود و اعلام آن در قالب یک عدد یا گزینه مشخص، نظر خود را به‌صورت توزیع شده و با انعطاف بالا از طریق تخصیص وزن یا باور به گزینه‌های مختلف بر روی مقیاس ارزیابی اعلام نماید (Murphy, 2000). این رویکرد برای استخراج دانش بسیار مناسب است و نیاز به تجمیع شواهد و انجام قضاوت توسط ارزیاب و در نهایت ایجاد خطا را کاهش می‌دهد. به‌عنوان‌مثال در این ساختار، ارزیاب می‌تواند میزان باور خود را بر روي طيف نظرسنجی توزیع نماید و به‌عبارت‌دیگر چند گزینه با میزان باورهاي متفاوت را انتخاب نماید. ازاین‌رو خطاهاي معمول مرتبط با ارزیاب (سخت‌گيري، سهل‌گيري، تمایل به مرکز) می‌تواند با ایجاد امکان انتخاب چند گزینه از طیف نظرسنجي با سطوح باورهاي متفاوت، تا حد زیادی تعدیل شود. از طرف دیگر در فرآیند ارزیابی مواقع بسیاری وجود دارد که در آن ارزیاب به دلیل نقص دانش بین انتخاب دو یا چند گزینه بر روی مقیاس ارزیابی مردد است. اعلام نظر در ساختار توابع باور امکان مشخص‌کردن هر دو گزینه به‌عنوان پاسخ را فراهم می‌آورد. همچنین ارزیاب می‌تواند نظر خود را به‌صورت ناقص اعلام داشته و بخشی از کل جرم باور را به‌عنوان جهل کامل به کل مجموعه گزینه‌های پاسخ تخصیص دهد. بدین

ترتیب سطوح مختلف نقص دانش بین یقین کامل تا عدم آگاهی کامل متناسب با سطح دانش در دسترس قابل مدل‌سازی است.

جدول ۱ به‌صورت شماتیک مثال‌هایی از نحوهٔ پاسخ‌دهی مبتنی بر توابع باور را در یک مقیاس ارزیابی فرضی پنج-سطحی با سطوح متفاوتی از میزان اطمینان ارزیاب نشان می‌دهد.

جدول شمارهٔ یک- مثال هایی از انعطاف ساختار باور در دریافت نظر از ارزیاب

میزان اطمینان ارزیاب در پاسخ‌دهی	نحوهٔ اعلام نظر در ساختار باور
با اطمینان صددرصد گزینهٔ ۴ صحیح است.	۱ ۲ ۳ (۴) ۵ $m(4) = 1$
با اطمینان کامل گزینهٔ صحیح ۳ یا ۴ است.	۱ ۲ (۳ ۴) ۵ $m(3,4) = 1$
شواهد متناقضی دریافت شده و اطلاعات کامل نیست. ۱۰ درصد باور به عملکرد در سطح گزینهٔ ۱ و ۲ و ۷۰ درصد باور به عملکرد در سطح گزینهٔ ۴ وجود دارد.	(۱ ۲) ۳ (۴) ۵ $m(1,2) = 0.1$ $m(4) = 0.7$ $m(\theta) = 0.2$
ارزیاب اطلاعات بسیار محدودی دارد. با همان اطلاعات محدود ۱۰ درصد باور به عملکرد در سطح گزینهٔ ۴ وجود دارد.	۱ ۲ ۳ (۴) ۵ $m(4) = 0.1$ $m(\theta) = 0.9$
ارزیاب اطلاعات محدودی دارد و فقط با اطمینان ۱۰۰ درصد اظهار می‌دارد که عملکرد در سطح گزینهٔ ۵ نمی‌باشد.	(۱ ۲ ۳ ۴) ۵ $m(1,2,3,4) = 1$
اطلاعات پراکنده‌ای در دسترس است. ۳۰ درصد باور به گزینهٔ ۲، ۲۰ درصد باور به گزینهٔ ۳، ۵۰ درصد باور به گزینهٔ ۴	۱ (۲) (۳) (۴) ۵ $m(2) = 0.3$ $m(3) = 0.2$ $m(4) = 0.5$

با استفاده از این ابزار می‌توان شواهد غیرقطعی حاصل از منابع مختلف ارزیابی عملکرد را با هر الگویی به‌راحتی و با استفاده از قوانین تجمیع شواهد، جمع‌بندی نمود. قابلیت تبیین میزان تعارض در بین شواهد و استفاده از آن در تصمیم‌گیری‌ها نیز از دیگر مزایای این ابزار است.

۲،۱،۲ مبانی تئوری شواهد

تئوری شواهد[1] که به آن تئوری دمستر - شیفر[2] یا تئوری توابع باور[3] نیز یاد می‌شود، بر اساس کار دمستر (۱۹۶۷) پایه‌گذاری و توسط شیفر (۱۹۷۶) توسعه داده شد. این تئوری ابزاری قوی جهت مدل‌سازی و ترکیب اطلاعات توأم با عدم قطعیت و تصمیم‌گیری بر اساس آن‌ها است. ساختار باورها اساس تئوري شواهد بوده و اساساً جهت مدل‌سازی ارزیابی‌های ذهني توأم با عدم قطعیت طراحي شده است، اما قابلیت تطبیق با داده‌هاي كمي، فازي و كيفي را نيز دارا است (Deng et al., 2011; Tang et al., 2014). در این تئوری هیچ الزامی وجود ندارد که باور تخصیص نیافته به یک گزاره[4] به نقیض آن تخصیص یابد. از طرف دیگر با استفاده از این تئوری می‌توان باور را علاوه بر گزاره‌های انفرادی به زیرمجموعه‌هایی از گزاره‌ها تخصیص داد بدون آنکه هیچگونه فرضی در مورد باور عناصر انفرادی زیرمجموعه‌ها در نظر گرفته شود (Sentz & Ferson, 2002). ویژگی‌های یادشده امکان مدل‌سازی مسأله را متناسب با سطح دانش در دسترس فراهم می‌آورد. یکی دیگر از جنبه‌های مهم تئوری شواهد امکان ترکیب شواهد حاصل از منابع مختلف و بررسی تعارض در بین آن‌هاست (Wang et al., 2019). این تئوری به جهت دارابودن ویژگی‌های ذکرشده و انعطاف مناسب در بسیاری از حوزه‌ها مانند صنعت نرم‌افزار(Paksoy & Göktürk, 2011) تحلیل ریسک(Deng et al., 2011; Taroun & Yang, 2011)،

[1] Evidence theory
[2] Dempster-shafer theory
[3] Belief function theory
[4] Proposition

تشخیص عیب (Jiang et al., 2016; Oukhellou et al., 2010; Qiang & Deng, 2022a; Vasu et al., 2015)، یادگیری عمیق (Tong et al., 2021) پیش‌بینی (Baraldi et al., 2015; Min et al., 2015; Xiao et al., 2012) و تصمیم‌گیری (AbuDahab et al., 2016; Fu & Yang, 2012) مورد استفاده قرار گرفته است. در ادامه مفاهیم و عناصر مرتبط با کاربرد تئوري شواهد به‌اختصار توضیح داده می‌شود.

- **چارچوب تمایز[1] (Θ)**

چارچوب تمایز دربرگیرندهٔ تمام گزاره‌های ممکن و منحصربه‌فردی است که منابع اطلاعاتی در مورد آن‌ها و ترکیبات مختلف آن‌ها، شواهد ارائه می‌نمایند. مجموعهٔ توانی 2^{Θ} که دربرگیرندهٔ تمام زیرمجموعه‌هاي ممکن Θ است، بدنهٔ شواهد نامیده می‌شود.

- **تخصیص باور پایه[2]**

یکی از مفاهیم کلیدی در تئوری شواهد تخصیص باور پایه یا توابع جرم باور[3] است. مقدار جرم باور زیرمجموعهٔ A که به‌صورت $m(A)$ نشان داده می‌شود، بیان‌کنندهٔ بخشی از باور کل است که از فرضیهٔ A پشتیبانی می‌کند و هیچ اطلاعات بیشتری در مورد تخصیص باور به زیرمجموعه‌های A در اختیار قرار نمی‌دهد. هر زیرمجموعهٔ $A \subset \Omega$ به‌نحوی‌که $m(A) > 0$ یک عنصر کانونی m نام دارد و جرم تخصیص‌یافته به کل چارچوب تمایز یا $m(\theta)$، جهل[4] نامیده می‌شود. تابع تخصیص باور پایه (m) خواص زیر را دارد:

$$m : 2^{\Theta} \to [0,1] \quad (۱)$$

$$m(\varnothing) = 0 \quad (۲)$$

$$\sum_{X \subset \theta} m(X) = 1 \quad (۳)$$

$$m(X) + m(\bar{X}) \le 1 \quad (۴)$$

- **ترکیب شواهد**

در تئوری شواهد بر اساس کیفیت و قابلیت اطمینان منابع فراهم آورنده شواهد، نحوهٔ گردآوری شواهد و رویکرد مواجهه با تعارض، قوانین ترکیب متعددی ارائه شده است (Awogbami et al., 2018; Lefevre et al., 2002; Lefèvre & Elouedi, 2013; Liu et al., 2011; Sentz & Ferson, 2002; Smets, 2007; Wang et al., 2019). در این مقاله قانون دمستر[5] یا ترکیب عطفی نرمال مورد استفاده قرار می‌گیرد که به‌صورت زیر تعریف می‌شود:

درصورتی‌که m1 و m2 را به‌عنوان تخصیص باور پایه از دو منبع متفاوت ۱ و ۲ در نظر بگیریم، آنگاه قانون ترکیب دمستر در ارتباط با گزاره A به شرح زیر تعریف می‌شود:

(5)

$$m_1 \oplus m_2(A) = \frac{\sum_{U,V|U \cap V = A} m_1(U).m_2(V)}{1 - \sum_{X,Y|X \cap Y = \varnothing} m_1(X).m_2(Y)}$$

که A در آن □□.

عبارت موجود در مخرج رابطهٔ فوق، عامل نرمال‌سازی است که معادل با ترکیب عطفی تمام شواهدی است که در تضاد[6] با هم نیستند.

قانون ترکیب دمستر (و دو قانون ذکر شده دیگر) داری ویژگی‌های جابه‌جایی و انجمنی بوده و به همین دلیل نسبت به اغلب قوانین ترکیب دیگر کاربرد بیشتری دارد (Dezert et al., 2012; Hooft et al., 2006; Zadeh, 1986). اصلی‌ترین ایراد وارد بر این قانون عدم توانایی آن در مواجهه با تعارض در شواهد[7] است. در این مقاله یکی از روش‌های متداول مواجهه با شواهد متعارض، تحت عنوان روش ضریب تعدیل مورد استفاده قرار می‌گیرد که در ادامه توضیح داده می‌شود.

- **روش ضریب تعدیل**

تعارض شواهد در وضعیتی ایجاد می‌شود که یک منبع قویاً از گزاره‌ای پشتیبانی نماید درحالی‌که منبع دیگر قویاً از گزاره‌ای ناسازگار با آن پشتیبانی کند. در کاربردهای تئوری شواهد با توجه به احتمال بروز تعارض در بین توابع باور و تأثیر آن بر کیفیت نتایج، نیاز است تا راهکاری مناسب جهت مواجهه با تعارض مورد استفاده قرار گیرد (Klein & Colot, 2011; Lefèvre & Elouedi, 2013; Schubert, 2011; Wang et al., 2019). یکی از ساده‌ترین و کاراترین روش‌های مواجهه با تعارض در تئوری شواهد که توسط شیفر ارائه شده است، روش ضریب تعدیل است و به‌صورت زیر تعریف می‌شود(Ha-Duong, 2008):

اگر m_j تابع جرم باور بر اساس اطلاعات منبع S_j و α_j میزان اطمینان به آن منبع باشد آنگاه تعدیل‌شده تابع m_j به‌صورت زیر تعریف می‌شود.

$$\begin{cases} m_{\alpha_j,j}(A) = \alpha_j m_j(A) \\ m_{\alpha_j,j}(\Theta) = 1 - \alpha_j + \alpha_j m_j(\Theta) \end{cases} \quad \forall A \subset \Theta \quad (۶)$$

[1] Frame of discernment
[2] Basic belief assignment
[3] Mass function
[4] Ignorance
[5] Dempster's rule of combination (DRC)
[6] conflict
[7] conflict

- **تصمیم‌گیری بر اساس توابع جرم باور**

ورودی‌ها و خروجی‌ها در تئوری شواهد بر اساس ساختار توابع جرم باور و یا تخصیص باورهای پایه بوده و اتخاذ تصمیم بر اساس آن‌ها نیازمند استفاده از معیارهایی است که امکان مقایسهٔ گزاره‌ها را فراهم می‌آورند (Wilson, 1993). یکی از معیارهای مناسب در این زمینه معیار احتمال پیگنیستیک[1] است که به‌صورت زیر تعریف می‌شود:

اگر m یک تخصیص باور پایه بر روی Ω باشد آنگاه معیارهای متداول در این حوزه به شرح ذیل هستند:

(۷)

$$BetP_m(\omega)=\sum_{A\subseteq\Omega,\omega\in A}\frac{1}{|A|}\frac{m(A)}{1-m(\varnothing)},m(\varnothing)\neq 1$$

که در آن $|A|$ بیانگر اندازهٔ مجموعه A و $\omega\in\Omega$ است(Smets, 2005; Wilson, 1993).

هرچه مقدار معیار ذکرشده بیشتر باشد گزارهٔ مورد بررسی برای انتخاب، گزینهٔ مطلوب‌تری است.

۲،۲ مطالعات مرتبط با افزایش کیفیت داده‌های ارزیابی

تلاش برای بهبود کیفیت ارزیابی‌ها سابقهٔ طولانی در علوم مدیریت دارد و بررسی تأثیر مقیاس‌های ارزیابی بر روی نظر ارزیاب و نتایج ارزیابی به‌طور گسترده در مطالعات مورد بررسی قرار گرفته است (Hedge & Teachout, 2000; Jelley et al., 2001; Levy & Williams, 2004; K. R. Murphy & Cleveland, 1995; Wagner & Goffin, 1997). در مطالعات متعدد به نقص سیستم نظرسنجي معمول، که در آن نظرات افراد از طریق مشخص‌کردن یک امتیاز بر روي مقیاس لیکرت دریافت می‌شود و نیز دشوار بودن پاسخ-گویي به سؤالات ذهني و کیفي در قالب یک عدد دقیق اشاره شده (Arbaiy & Suradi, 2007; Golman & Bhatia, 2012; Levy & Williams, 2004; Manoharan et al., 2011) و به لزوم مطالعه در زمینهٔ مقیاس‌هاي ارزیابي و استفاده از مقیاس‌هایي متناسب با هدف مطالعه تأکید شده است (Manoharan et al., 2011; Nowack & Mashihi, 2012). در بررسی مطالعات در این حوزه دو رویکرد در جهت بهبود کیفیت داده‌های فرآیند ارزیابی عملکرد قابل شناسایی است.

رویکرد اول که بیشتر در بین مطالعات انجام‌شده با زمینهٔ علوم انسانی و روان‌شناسی مشاهده می‌شود، در قالب غنی‌سازی مقیاس ارزیابی و یا ارائهٔ آموزش به ارزیاب در جهت ایجاد استاندارد و مرجع مشترک ارزیابی تعریف می‌شود. این راهکارها با هدف کمک به ارزیاب در جهت انجام قضاوت صحیح‌تر و عینی‌تر و در نهایت افزایش کیفیت و قابلیت اطمینان نتایج توسعه یافته‌اند (Hoffman et al., 2012). ازجمله مهم‌ترین راهکارهایی که با این رویکرد توسعه یافته‌اند می‌توان به مقیاس رتبه‌بندی رفتاری، مقیاس چارچوب مرجع ارزیابی و راهکار آموزش چارچوب مرجع ارزیابی اشاره نمود. در مقیاس رتبه‌بندی رفتاری، هر کدام از گزینه‌های پاسخ بر روی مقیاس ارزیابی با استفاده از یک مثال و نمونه از رفتار متناظر توضیح داده می‌شود. در مطالعات متعدد بهبودهای حاصل از این مقیاس ارزیابی گزارش شده است(Borman et al., 2001a; Jelley et al., 2001). باوجوداین طراحی این مقیاس و تعیین امتیازهای دقیق متناظر با هر الگوی رفتاری در آن دشوار است(John Bernardin et al., 2016). از طرفی همان‌طور که اشاره شد ارزیاب‌ها ممکن است رفتارهای متفاوت و متناقضی از فرد ارزیابی‌شونده مشاهده نمایند و به‌طورکلی ممکن است وضعیتی ایجاد شود که در آن ارزیاب به دلیل نقص دانش، بین انتخاب دو یا چند گزینه پاسخ مردد باشد. مقیاس رتبه‌بندی رفتاری فاقد قابلیت لازم جهت مدل‌سازی چنین وضعیتی است. در مقیاس چارچوب مرجع ارزیابی، توضیحات و اطلاعات تکمیلی، قبل از بیان سؤالات ارزیابی در هر بخش ارائه می‌شوند (Hoffman et al., 2012). این مقیاس نیز با وجود دارابودن برخی مزایا در بهبود کیفیت داده‌های ارزیابی، فاقد قابلیت مدل‌سازی عدم قطعیت ناشی از نبود دانش است.

راهکار آموزش چارچوب مرجع ارزیابی نیز مشابه با دو راهکار معرفی شده است با این تفاوت که در آن به‌جای بیان توضیحات تکمیلی بر روی پرسش‌نامه، آموزش‌های طراحی‌شده در این زمینه با بیان مثال‌های متعدد به ارزیاب‌ها ارائه می‌شود(Hoffman et al., 2012) . مشابه موارد قبل، ارائه این آموزش‌ها به ارزیاب در بهبود کیفیت داده‌های ارزیابی مؤثر است. ازجمله نقاط ضعف این راهکار، هزینه‌بر و زمان‌بر بودن آن است که با افزایش تعداد ارزیاب‌ها این مسأله تشدید می‌گردد. این راهکار نیز مانند موارد قبل فاقد قابلیت مدل‌سازی عدم قطعیت ناشی از نقص دانش است.

با رویکردی متفاوت، در برخی از مطالعات که مطالعهٔ حاضر نیز در این دسته قرار می‌گیرد، با اشاره به وجود عدم قطعیت در فرآیند ارزیابی عملکرد، به لزوم کاربرد ابزار مناسب جهت مدل‌سازی و مواجهه با عدم قطعیت در جهت افزایش کیفیت داده‌های ارزیابی تأکید شده است (Arbaiy & Suradi, 2007; Espinilla et al., 2013; Golec & Kahya, 2007; Moon et al., 2010). اهم مطالعات انجام‌شده با این رویکرد در فرآیند ارزیابی عملکرد کارکنان، به کاربرد ابزارهای فازی در طراحی مقیاس‌های ارزیابی مرتبط است. در این راستا، مقیاس ارزیابی مبتنی بر عبارات زبانی فازی یا مقیاس تبدیل فازی[2] متداول‌ترین نوع کاربرد ابزارهای فازی در فرآیند ارزیابی عملکرد کارکنان است. مبنای اساسی طراحی این مقیاس بر این حقیقت قرار دارد که استفاده از عبارات زبانی به‌جای گزینه‌های عددی، به فرآیند قضاوت‌های انسانی نزدیک‌تر بوده و اعلام نظر ارزیاب را تسهیل می‌نماید(Arbaiy & Suradi, 2007; De La Rosa

[1] Pignistic Probability

[2] Fuzzy conversion scale

(De Sáa et al., 2013; Zadeh, 1975). این مقیاس را می‌توان نوع خاصی از مقیاس لیکرت دانست که در آن سؤالات و گزینه‌های ارزیابی در قالب متغیرها و عبارات زبانی فازی بیان می‌شوند. در این مقیاس ارزیاب نظر خود را با مشخص‌کردن نزدیک‌ترین عبارت زبانی به قضاوت خود اعلام می‌کند. ازآنجاکه خروجی این مقیاس به اعداد فازی تبدیل می‌شود و اعداد فازی خود توابع حقیقی هستند می‌توان اطلاعات محتوای آن‌ها را با استفاده از روش‌های ریاضی بررسی و استخراج نمود. این روش به جهت مدل‌سازی عدم قطعیت ناشی از مفاهیم ذهنی و فازی و سهولت تفهیم آن برای ارزیاب کاربردهای گسترده‌ای داشته است اما مشابه روش‌های گذشته فاقد قابلیت مدل‌سازی نقص دانش است.

مقیاس کاملاً متفاوت دیگر که مبتنی بر مفاهیم فازی توسعه‌یافته است، مقیاس رتبه‌بندی فازی نامیده می‌شود. این مقیاس ارزیابی با عناوین قالب پاسخ آزاد[1] و ارزش‌گذاری فازی[2] توسط هسکت[3] و همکاران ارائه شده است (Hesketh, 1988). در این مقیاس، اعلام نظر ارزیاب محدود به یک لیست ثابت و از پیش تعیین شده مقادیر فازی نمی‌باشد. ازاین‌رو در مقیاس رتبه‌بندی فازی با توجه به آزادی در ارائهٔ ارزش‌گذاری‌ها و پاسخ‌ها، تنوع و تعدد ارزیابی‌ها بسیار بیشتر است. علی‌رغم قابلیت‌های این مقیاس ارزیابی در دریافت بهتر نظر از ارزیاب و مدل‌سازی عدم قطعیت ناشی از نقص دانش در آن، کاربرد آن با محدودیت‌های اساسی مواجه است. ازجملهٔ این محدودیت‌ها می‌توان به نبود یا کمبود روش‌های آماری تحلیل داده‌های حاصل از آن، نیاز به آموزش ارزیاب‌ها جهت استفاده از آن و صرف زمان نسبتاً زیاد در تکمیل آن اشاره نمود(Lubiano et al., 2016a).

جدول شمارهٔ دو- مطالعات مرتبط با افزایش کیفیت ارزیابی‌ها

رویکرد	تکنیک	مزایا	معایب	مقالات مرتبط
()	غنی‌سازی مقیاس ارزیابی (مثال: مقیاس رتبه‌بندی رفتاری و مقیاس ارزیابی چارچوب مرجع)	ایجاد درک مشترک بین ارزیاب‌های مختلف در مورد مقیاس ارزیابی	دشوار بودن تعریف نمونه‌های رفتاری اثربخش و غیر اثربخش در حوزه‌های مختلف، نپرداختن به عدم قطعیت ناشی از نقص دانش	(Borman et al., 2001b; Hoffman et al., 2012; Murphy, 2008b)
	آموزش چارچوب مرجع ارزیابی	بهبود کیفیت داده‌های ارزیابی	هزینه‌بر و زمان‌بر بودن، عدم ارائهٔ راهکار جهت مدل‌سازی عدم قطعیت ناشی از نقص دانش	Woehr & Huffcutt, 1994b;) (Woehr, 2008
کاربرد ابزارهای مدل‌سازی و مواجهه با عدم قطعیت	مقیاس عبارات زبانی فازی	سادگی کاربرد مدل‌سازی عدم قطعیت ناشی از مفاهیم فازی	نبود امکان مدل‌سازی نبود قطعیت ناشی از نقص دانش	Chang et al., 2007; Esen et) al., 2016; Jamsandekar & Mudholkar, 2013; Macwan & Sajja, 2013; Meenakshi, 2012; (Shaout et al., 2014
	مقیاس ارزیابی فازی	قابلیت دریافت اطلاعات بیشتر قابلیت مدل‌سازی عدم قطعیت ناشی از نقص دانش	دشواری اجرا - کمبود روش‌های تحلیل داده‌های آن	(De La Rosa De Sáa et al., 2013; Lubiano et al., 2016a; Lubiano et al., 2016b)
	مقیاس ارزیابی مبتنی بر ساختار باور	مزایای پیش‌بینی‌شده: تسهیل فرآیند اعلام نظر، افزایش کیفیت نتایج ارزیابی با توجه به دارا بودن قابلیت مدل‌سازی همزمان عدم قطعیت ناشی از نقص دانش و عدم قطعیت تصادفی	نیاز به محاسبات نسبتاً زیاد در تجمیع داده‌ها	مقالهٔ حاضر

[1] Freeresponse format
[2] Fuzzy valuation
[3] Hesketh

همان‌طور که در جدول ۲ آمده، تاکنون مطالعات متعددی در جهت بهبود کیفیت ارزیابی‌ها انجام شده است که هر کدام مزایا و معایب خاص خود را دارد؛ در جدول مذکور، در رویکرد اول، به‌منظور کاهش عدم‌قطعیت پاسخ‌ها، به‌جای استفاده از ابزار متناسب «مواجهۀ با عدم قطعیت»، به شفاف‌سازی و استانداردسازی امتیازات و مقیاس ارزیابی پرداخته شده است؛ این رویکرد قابلیت استفاده در همۀ نظرسنجی‌ها را نداشته و نقطه ضعف عمدۀ آن هزینه و زمان زیاد فرآیند ارزیابی است. در رویکرد دوم، مطابق آن‌چه در جدول آمده، دو تکنیک مورد استفاده قرار گرفته است؛ در روش «عبارات زبانی فازی» قابلیت مدل‌سازی «عدم قطعیت ناشی از نقص دانش» وجود ندارد و این مورد نقطه ضعف اساسی تکنیک مذکور است؛ هم‌چنین نقطه ضعف تکنیک «مقیاس ارزیابی فازی» را می‌توان در دشواری پاسخ‌گویی دانست. دراین‌بین راهکار «ساختار توابع باور» که علاوه بر انعطاف و تسهیل در اعلام نظر ارزیاب، قابلیت مدل‌سازی هم‌زمان «عدم قطعیت ناشی از نقص دانش» و «عدم قطعیت تصادفی» را داراست، تاکنون به‌عنوان سنجۀ ارزیابی مورد بررسی قرار نگرفته است؛ ازاین‌رو در پژوهش حاضر امکان استفاده از مقیاس ارزیابی مبتنی بر ساختار باور و مزایای متصور از کاربرد آن مورد بررسی قرار می‌گیرد.

۳ اهداف و فرضیات پژوهش

همان‌طور که پیش‌تر بیان شد راهکار اعلام نظر منطبق بر ساختار باور با قابلیت دریافت نظرات به‌صورت توزیع شده، شرایط اعلام نظر متناسب با سطح دانش ارزیاب را فراهم آورده و از این طریق، می‌تواند ضمن تسهیل اعلام نظر ارزیاب، امکان دریافت دقیق‌تر نظرات را فراهم آورد. در فرآیند ارزیابی مبتنی بر پرسش‌نامه، مواردی وجود دارد که ارزیاب در مورد شاخص ارزیابی اطلاعات کافی در اختیار ندارد. اعتقاد ما بر این است که مقیاس ارزیابی مبتنی بر ساختار باور در این شرایط نسبت به سایر مقیاس‌های متداول ارزیابی، شامل مقیاس ارزیابی لیکرت، مقیاس ارزیابی مبتنی بر عبارات فازی و مقیاس ارزیابی مشابهت بصری، می‌تواند کارایی بهتری چه از نظر کیفیت نتایج و چه از نظر راحتی ارزیاب داشته باشد؛ ازاین‌رو اهداف این پژوهش به‌صورت زیر تعیین می‌شود:

- بررسی مقایسه‌ای عملکرد سنجه‌های چهارگانۀ عملکردی در شاخص‌های ارزیابی ساده و پیچیده.
- بررسی مقبولیت نتایج حاصل از مقیاس‌های مختلف ارزیابی.
- بررسی سهولت کار با مقیاس‌های مختلف ارزیابی

لذا فرضیه‌های پژوهش به شرح زیر است:

- **فرضیۀ ۱:** عملکرد مقیاس‌های لیکرت، عبارات فازی، مشابهت بصری و ساختار باور در سنجش شاخص‌های ساده و بدون پیچیدگی متفاوت است.
- **فرضیۀ ۲:** عملکرد مقیاس‌های لیکرت، عبارات فازی، مشابهت بصری و ساختار باور در سنجش شاخص‌هایی که اطلاعات کمتری در مورد آن‌ها وجود دارد، متفاوت است.
- **فرضیۀ ۳:** بین مقبولیت نتایج حاصل از مقیاس‌های مختلف ارزیابی، تفاوت معنادار وجود دارد.
- **فرضیۀ ٤:** بین سهولت کار با مقیاس‌های مختلف ارزیابی، تفاوت معنادار وجود دارد.

۴ روش‌شناسی پژوهش

شالوده و جهان‌بینی پژوهش حاضر اثبات‌گرایی است و از رویکرد استقرایی به‌منظور انجام تحقیق استفاده شده است. در این تحقیق که با هدفی کاربردی انجام گرفته است با بهره‌گیری از استراتژی پیمایشی، ابزار پرسش‌نامه به‌منظور بررسی فرضیات به روش کمی برگزیده شد.

٤,۱ نمونۀ آماری و ابزار جمع‌آوری داده

همان‌طور که پیش‌تر عنوان شد، در پژوهش حاضر، هدف، بررسی مقایسه‌ای «سنجه‌های ارزیابی» است؛ به‌عبارت‌دیگر محتوا و ماهیت مفهوم مورد ارزیابی، موضوع پژوهش نبوده است؛ ازاین‌رو جامعۀ آماری دانشگاهیان (با توجه به سهولت دسترسی به این جامعه)، به‌منظور دست‌یابی به اهداف تحقیق مورد استفاده قرار گرفت؛ بدین ترتیب با استفاده از شاخص‌های متداول ارزیابی اساتید که در سطح دانشگاه‌های کشور مورد استفاده قرار می‌گیرد دو پرسش‌نامه با سؤالات محدود تهیه شد و یکی از کلاس‌های دورۀ کارشناسی گروه فنی مهندسی دانشگاه پیام نور به‌عنوان مرجع گردآوری داده مورد استفاده قرار گرفت. در مرحلۀ اول (به‌منظور بررسی فرضیۀ اول پژوهش) پرسش‌نامه‌ای متشکل از پنج سؤال متداول جهت ارزیابی یک استاد، توسط پنج دانشجو طراحی شد. سؤالات به‌گونه‌ای انتخاب شدند که اعلام نظر توسط دانشجویانی که در کلاس درس حاضر بودند به‌راحتی ممکن باشد (جدول ۳).

جدول شمارۀ سه- سؤالات طرح‌شده در پرسش‌نامۀ اول

سؤالات پرسش‌نامۀ اول
شروع به‌موقع کلاس درس
برگزاری مؤثر کلاس در زمان استاندارد تعیین‌شدۀ آن
پوشش کامل سرفصل آموزشی درس
ارائۀ طرح درس در جلسۀ اول کلاس
اعلام به‌موقع نتایج ارزش‌یابی‌ها

این سؤالات در قالب چهار پرسش‌نامهٔ جداگانه که سؤالات آن‌ها کاملاً مشابه بوده و تنها در مقیاس پاسخ‌گویی با یکدیگر تفاوت دارند، شامل مقیاس ارزیابی عددی لیکرت، مقیاس ارزیابی شباهت بصری، مقیاس ارزیابی عبارات زبانی فازی و مقیاس ارزیابی مبتنی بر ساختار باور پیاده‌سازی گردید. در پرسش‌نامهٔ طراحی‌شده مبتنی بر مقیاس لیکرت از طیف ۵ گزینه‌ای عددی (گزینه ۱ تا ۵) استفاده شد. در مقیاس شباهت بصری ارزیابی‌ها با استفاده از طیف ارزیابی از یک تا ۱۰۰ (گزینهٔ ۱ بسیار ضعیف، گزینهٔ ۱۰۰ بسیار خوب) انجام می‌شود. در مقیاس عبارات زبانی فازی، همان‌طور که در جدول ۴ آمده است، از پنج گزینهٔ بسیار ضعیف، ضعیف، متوسط، خوب و بسیار خوب، متناظر با مقیاس تبدیل مبتنی بر اعداد مثلثی، استفاده شد. در روش پاسخ‌گویی مبتنی بر ساختار باور نیز از گزینه‌های مشابه با مقیاس لیکرت در طیف ارزیابی استفاده گردید.

جدول شمارهٔ چهار- معادل فازی عبارات زبانی استفاده شده

عبارت زبانی	کد اختصاری	عدد مثلثی معادل
بسیار خوب	VG	(4,5,5)
خوب	G	(3,4,5)
متوسط	M	(2,3,4)
ضعیف	W	(1,2,3)
بسیار ضعیف	VW	(1,1,2)

با توجه به اینکه هدف از اجرای میدانی این پژوهش، مقایسهٔ قابلیت ابزارهای ارزیابی بوده و نیز در این پژوهش هر ارزیاب در هر ارزیابی می‌بایست چهار بار یک پرسش‌نامه (با مقیاس‌های چهارگانه) را تکمیل نماید، سعی شد پرسش‌نامه تا حد امکان خلاصه طراحی گردد. مورس[1] و ساریس[2] (۱۹۹۰) نشان داده‌اند که پس از گذشت ۲۰ دقیقه ارزیاب‌ها پاسخ‌های قبلی خود را فراموش می‌کنند و تأثیری بر تکمیل پرسش‌نامه بعدی نخواهد داشت(Saris & van Meurs, 1990) . لذا فاصلهٔ زمانی بین تکمیل پرسش‌نامه‌ها حداقل نیم ساعت تعیین گردید که در برخی موارد به چند روز نیز افزایش می‌یافت. پرسش‌نامه‌ها به صورت الکترونیکی و با توضیحات کافی در فواصل زمانی مشخص در اختیار دانشجویان قرار گرفت.

در مرحلهٔ بعد (به‌منظور بررسی فرضیهٔ دوم پژوهش) یک پرسش‌نامه که ازنظر ساختاری با پرسش‌نامهٔ قبلی کاملاً مشابه بود با سه سؤال جدید طراحی شد. سؤالات این پرسش‌نامه مفهومی‌تر بوده و پاسخ‌گویی به آن‌ها به‌راحتی و تنها با حضور در کلاس امکان‌پذیر نبود و نیازمند مشارکت و تعامل بیشتر دانشجو با استاد بود (جدول ۵). در انتهای هر پرسش‌نامه (به‌منظور بررسی فرضیهٔ چهارم) یک سؤال در خصوص راحتی کار با مقیاس ارزیابی پرسش‌نامه وجود داشت که در این سؤال، ارزیاب امتیازی بین ۱ تا ۵ در مورد میزان راحتی کار با مقیاس ارزیابی ارائه می‌نمود.

جدول شمارهٔ پنج- سؤالات طرح‌شده در پرسش‌نامهٔ دوم

سؤالات پرسش‌نامهٔ دوم
رعایت اخلاق و شئون حرفه‌ای
ایجاد انگیزهٔ یادگیری دانش و مهارت در دانشجویان
همکاری مسؤولانه در ارائهٔ مشاوره‌های آموزشی و تحصیلی به دانشجویان

به‌منظور بررسی فرضیهٔ سوم پژوهش (سنجش مقبولیت نتایج به‌دست‌آمده از مراحل یک و دو) پرسش‌نامهٔ دیگری طراحی شد که در آن، سؤالات، همان سؤالات دو مرحلهٔ قبل بوده و گزینه‌های پاسخ به هر سؤال، تجمیع نتایج حاصل از نظرسنجی ارزیاب‌ها (در مراحل گذشته) به تفکیک چهار مقیاس ارزیابی تعیین گردید. تجمیع داده‌های ارزیابی به شرح جدول ۶ انجام گرفت. در این مرحله از نظرسنجی، پرسش‌نامه‌ها در اختیار دانشجویان و مدیر گروه (۱۵ نفر) قرار داده شد تا مورد قبول‌ترین نتیجه به‌دست‌آمده را مشخص نمایند. این گزینه بالاترین رتبهٔ معادل عدد ۴ را کسب نموده و به سایر گزینه‌ها بر اساس فاصله از گزینهٔ انتخابی، رتبه‌های ۳، ۲ و ۱ تخصیص داده شد.

[1] Meurs

[2] Saris

جدول شمارهٔ شش- نحوهٔ تجمیع داده‌های مرتبط با هر معیار ارزیابی

مقیاس ارزیابی	روش تجمیع و استخراج نتیجه
مقیاس ارزیابی لیکرت	میانگین ساده
مقیاس ارزیابی عبارات فازی	میانگین فازی/روش سنتروید[1]
مقیاس شباهت بصری	میانگین ساده
ارزیابی در قالب توابع باور	قانون دمستر/ معیار احتمال پیگنیستیک

۴،۲ ابزار تجزیه‌وتحلیل

جهت بررسی معنادار بودن تفاوت میان داده‌های به‌دست‌آمده از مقیاس‌های مختلف ارزیابی (فرضیات یک و دو)، با توجه به بررسی پیش‌فرض داده‌ها از آزمون‌های پارامتریک (آنالیز واریانس یک‌طرفه با مقادیر تکراری (گروه‌های همبسته)) و ناپارامتریک متناسب (فریدمن) استفاده شد. در این تحلیل هدف اصلی، بررسی نقش عامل «مقیاس‌های ارزیابی»، بر نظرات دریافت شده از ارزیاب‌ها است.

جهت بررسی معنادار بودن تفاوت میان مقبولیت نتایج حاصل از مدل‌های ارزیابی و اولویت‌بندی آن‌ها (فرضیهٔ سوم)، به جهت نرمال نبودن داده‌ها از آزمون رتبه‌بندی فریدمن استفاده شد.

به‌منظور بررسی متغیر «سهولت کار با مقیاس‌های ارزیابی» در فرضیهٔ چهارم به دلیل نرمال نبودن داده‌ها آزمون رتبه‌بندی فریدمن انتخاب شد.

۵ یافته‌های پژوهش

۵،۱ بررسی فرضیهٔ اول: تفاوت نتایج به‌دست‌آمده از مقیاس‌های ارزیابی در سؤالات ساده

نتایج حاصل از ارزیابی استاد توسط پنج دانشجو کلاس، با استفاده از پرسش‌نامهٔ اول در جدول ۷ آمده است. در این جدول داده‌های خام حاصل از پرسشنامه ارزیابی و داده‌های هم‌مقیاس شده و استخراج‌شده از مقیاس‌های مختلف ذکر شده است. استخراج داده‌های هم‌مقیاس شده در ستون عبارات فازی بر اساس روش سنتروید انجام شده است. داده‌های خام در ستون عبارات فازی بر اساس کد اختصاری ارائه‌شده در جدول ۴ بیان شده‌اند. نتایج ستون مشابهت بصری از تبدیل داده‌های به‌دست‌آمده در مقیاس صفر تا صد به مقیاس یک تا پنج حاصل شده است. در ستون ساختار باور، نتایج با معیار احتمال پیگنیستیک به اعداد حقیقی تبدیل شده است.

جدول شمارهٔ هفت- نتایج ارزیابی با استفاده از پرسش‌نامهٔ اول

سؤال	نوع داده	مقیاس لیکرت	عبارات فازی	مشابهت بصری	ساختار باور
۱	داده‌های خام	5,4,4,5,4	VG, VG, M, G, VG	88,90,62,76,90	m{4}=1,m{5}=1,m{4}=1,m{45}=1,m{5}=1
	هم مقیاس شده	--	4.75,4.75,3,4,4.75	4.4,4.5,3.1,3.8,4.5	5,4.5,4,5,4
۲	داده‌های خام	4,4,5,3,4	G,G,VG,M,M	78,71,98,64,51	m{4}=1,m{34}=0.5 m{3}=0.5, m{5}=1,m{4}=1,m{34}=1
	هم مقیاس شده	--	4,4,4.75,3,3	3.9,3.55,4.9,3.2,2.55	3.5,4,5,3,4
۳	داده‌های خام	3,3,3,3,2	W,W,W,G,M	41,40,38,80,60	m{3}=1,m{34}=1,m{3}=1,m{3}=1, m{34}=0.5 m{3}=0.5

[1] Centroid

	هم مقیاس شده	--	2,2,2,4,3	2.05,2,1.9,4,3	3,3,3,3.5,3
٤	داده‌های	4,5,4,4,4	VG,G,G,G,G	90,82,66,78,76	m{4}=0.5 m{45}=0.5, m{34}=1, m{4}=1, m{5}=1, m{4}=0.5 m{45}=0.5
	هم مقیاس شده	--	4.75,4,4,4,4	4.5,4.1,3.3,3.9, 3.8	4,5,4,3.5,4
٥	داده‌های	5,4,5,5, 5	VG,VG,VG,VG, VG	94,86,98,98,90	m{5}=1, m{5}=0.5 m{45}=0.5, m{5}=1, m{4}=0.5 m{45}=0.5, m{5}=1
	هم مقیاس شده	--	4.75,4.75,4.75,4. 75,4.75	4.7,4.3,4.9,4.9, 4.5	5,4,5,5,5

از پیش‌فرض‌های مطرح در آنالیز واریانس، نرمال‌بودن داده‌های هر گروه است. اگرچه حساسیت نتایج در این روش نسبت به داده‌های غیرنرمال زیاد نیست اما پیش از آزمون مورد بررسی قرار گرفتند؛ آزمون گلموگروف-اسمیرونوف[1] نرمال بودن داده‌ها را تأیید نکرد؛ اما بررسی چولگی و کشیدگی[2] داده‌ها، قدرمطلق تمام این مقادیر را کمتر از دو نشان داد که به بیان لوین و روبین (۱۹۹۸) ویژگی داده‌های تقریباً نرمال را داراست؛ ازاین‌رو می‌توان از آنالیز واریانس یک‌طرفه با داده‌ها تکراری به بررسی فرضیه پرداخت؛ لازم به ذکر است با توجه به وابسته‌بودن آزمودنی‌ها، این شیوه در مقایسه با آنالیز واریانس با مشاهدات مستقل، دقت و توان بالاتر دارد.

کرویت[3] داده‌ها بر اساس تست موچلی[4] رد شد؛ چراکه معناداری آزمون مزبور رقم ۰,۰۰۰ را نشان می‌داد؛ بنابراین به‌منظور آزمون فرضیه به سایر شاخص‌ها رجوع شد؛ هر سه شاخص گرینهوس-گیسر، هیون-فیلد، و لور بند[5] با مقادیر ۰,۰٦٤، ۰,۰٦۲ و ۰,۰۷۹ فرضیهٔ برابری میانگین امتیازات حاصل از مقیاس‌های مختلف را تأیید کردند. این بدان معناست که در سؤالات ساده و فاقد پیچیدگی، تفاوت بین نتایج حاصل از مقیاس‌های ارزیابی، اهمیت چندانی ندارد.

۵,۲ بررسی فرضیهٔ دوم: تفاوت نتایج به‌دست‌آمده از مقیاس‌های ارزیابی در سؤالات پیچیده

نتایج حاصل از ارزیابی استاد توسط[5] دانشجوی کلاس، با استفاده از پرسش‌نامهٔ دوم که شامل سه سؤال است، در جدول ۸ آمده است. داده‌های این جدول نیز مطابق توضیحات ذکرشده در بخش ۴,۱ هم مقیاس شده است.

[1] Kolmogorov-Smirnov Test
[2] Skewness & Kurtosis
[3] Sphercity
[4] Mauchly's Test
[5] Greenhouse-Geisser, Huynh-Feldt, Lower-bound

جدول شمارهٔ هشت- نتایج ارزیابی با استفاده از پرسش‌نامهٔ دوم

سؤال	نوع داده	مقیاس لیکرت	عبارات فازی	مشابهت بصری	ساختار باور
۱	داده‌های خام	4,5,4,3,4	M,G,M,G,G	63,68,74,57,72	m{34}=1,m{34}=1,m{23}=1,m{34}=0.5 m{3}=0.5,m{34}=1
	هم مقیاس شده	--	3,4,3,4,4	3.15,3.4,3.7,2.85,3.6	3.5,3.5,2.5,3,3.5
۲	داده‌های خام	4,4,3,4,4	G,M,M,G,G	75,73,95,70,64	m{34}=0.5 m{4}=0.5,m{34}=0.5 m{3}=0.5,m{34}=0.5 m{3}=0.5,m{34}=1,m{3}=1
	هم مقیاس شده	--	4,3,3,4,4	3.75,3.65,4.75,3.5,3.2	4,3,3,3.5,3
۳	داده‌های خام	4,4,3,4,4	G,G,G,M,G	74,66,68,72,60	m{34}=0.5m{3}=0.5, m{34}=1, m{34}=0.5 m{3}=0.5,m{34}=0.5 m{3}=0.5,m{34}=1
	هم مقیاس شده	--	4,4,4,3,4	3.7,3.3,3.4,3.6,3	3,3.5,3,3,3.5

نظیر داده‌های فرضیهٔ اول، آزمون گلموگروف-اسمیرونوف نرمال بودن داده‌های مقیاس‌های مشابهت بصری و ساختار باور را تأیید نکرد. همچنین بررسی چولگی و کشیدگی داده‌ها، میزان کشیدگی داده‌های حاصل از مقیاس مشابهت بصری را برابر ۴,۲۴ نشان داد که به بیان لوین و روبین (۱۹۹۸) ویژگی داده‌های تقریباً نرمال را دارا نیست؛ ازاین‌رو از آزمون ناپارامتریک فریدمن به بررسی فرضیه پرداخته شد؛ سطح معناداری ۰,۰۱ فرضیهٔ برابری مقادیر حاصل از مقیاس‌های چهارگانه را رد کرد. این بدان معناست که در سؤالاتی که ارزیاب اطلاعات کاملی جهت اعلام نظر ندارد یا سؤالات طوری طراحی شده که امکان اعلام نظر قطعی را دشوار می‌سازد، مقیاس‌های مختلف ارزیابی نظرات متفاوتی را ارائه می‌کنند.

به‌منظور بررسی دقیق‌تر مطابق جدول ۹ مقایسهٔ دو به دوی نتایج حاصل از مقیاس‌ها با استفاده از آزمون ناپارامتریک ویلکاکسون[1] انجام گرفت:

سطوح معناداری جدول ۹ نشان می‌دهد تفاوت نتایج حاصل از مقیاس‌های چهارگانه، صرفاً ناشی از تفاوت مقیاس ساختار باور با دو مقیاس لیکرت و عبارات فازی است. بر اساس این تحلیل، تفاوت معناداری میان نتایج حاصل از مقیاس ساختار باور با مشابهت بصری ملاحظه نشد.

جدول شمارهٔ نه- سطوح معناداری مقایسهٔ دو به دوی نتایج حاصل از مقیاس‌ها در پرسش‌نامهٔ دوم

سنجه‌های مورد مقایسه	P. Valueدوسویه
لیکرت-عبارات فازی	۰,۲۶
لیکرت- مشابهت بصری	۰,۰۳۱
لیکرت- ساختار باور	۰,۰۰۳
عبارات فازی- مشابهت بصری	۰,۳۲

[1] Wilcoxon

عبارات فازی-ساختار باور	۰,۰۰۸
مشابهت بصری-ساختار باور	۰,۱۸

۵,۳ بررسی فرضیهٔ سوم: مقایسهٔ مقبولیت نتایج به‌دست‌آمده

در این مرحله، موردقبول‌ترین نتیجه از میان برآیندهای تجمیعی حاصل از هر یک از مقیاس‌های چهارگانه به تفکیک دو پرسش‌نامهٔ ساده و پیچیده مورد بررسی قرار گرفت. با توجه به نتایج حاصل از فرضیات اول و دوم مشخص شد صرفاً میان برآیندهای تجمیعی حاصل از پرسش‌نامه مربوط به سؤالات پیچیده، شاهد تفاوت معنادار هستیم ازاین‌رو آزمون مقبولیت نتایج، بر روی مقیاس‌های چهارگانهٔ حاصل از پرسش‌نامهٔ دوم انجام گرفت. در این مرحله نتایج تجمیع‌شدهٔ ارزیابی‌ها در پرسش‌نامهٔ دوم محاسبه شد که خروجی آن در جدول ۱۰ آمده است. در این جدول همچنین تفاوت میان نتایج تجمیع شدهٔ سایر مقیاس‌های ارزیابی و مقیاس ارزیابی مبتنی بر ساختار باور آمده است. همان‌طور که در این جداول نشان داده شده است، مجموع و میانگین تفاوت نتایج نهایی به‌دست‌آمده از ارزیابی‌ها در پرسش‌نامهٔ دوم که در آن سؤالات ارزیابی به‌صورت پیچیده طراحی شده، قابل توجه و تعیین‌کننده است.

جدول شمارهٔ ده- جمع‌بندی نتایج ارزیابی‌های پرسش‌نامهٔ دوم به تفکیک مقیاس‌های ارزیابی

سؤال	نتایج تجمیع شده				فاصلهٔ نتایج ساختار باور و سایر مقیاس‌ها		
	در مقیاس لیکرت	عبارات فازی	مشابهت بصری	ساختار باور	تفاوت با مقیاس لیکرت	تفاوت با عبارات زبانی فازی	تفاوت با مقیاس مشابهت بصری
۱	٤	۳,٦	۳,۳۹	۳	۱	۰,٦	۰,۳۹
۲	۳,۸	۳,٦	۳,۷۸	۳	۰,۸	۰,٦	۰,۷۸
۳	۳,۸	۳,۸	۳,۳۳	۳	۰,۸	۰,۸	۰,۳۳
مجموع تفاوت بین نتایج حاصل از ساختار باور و سایر مقیاس‌ها					۲,٦۰	۲,۰۰	۱,٤۹
میانگین تفاوت بین نتایج حاصل از ساختار باور و سایر مقیاس‌ها					۰,۸۷	۰,٦۷	۰,۵۰

در ادامهٔ کار بر اساس نتایج تجمیع داده‌های ارزیابی، پرسش-نامه‌ای مطابق توضیحات بند ۳,۲ تهیه و نظرسنجی مقبولیت نتایج انجام گرفت. بر اساس داده‌های گردآوری‌شده به تفکیک سؤالات پرسش‌نامهٔ دوم (سه سؤال)، آزمون فریدمن اجرا شد. این نتایج در جدول ۱۱ آمده است. نتایج نشان می‌دهد مقبولیت نتایج حاصل از مقیاس‌های مختلف ارزیابی به‌صورت معناداری متفاوت است.

همان‌طور که در جدول ۱۱ آمده است، بین مقبولیت خروجی‌های مختلف به‌دست‌آمده از مقیاس‌های ارزیابی، تفاوت معنادار وجود دارد. نتایجی که در سطر مربوط به میانگین رتبه‌های این جدول آمده است، نشان می‌دهد خروجی ارزیابی‌های انجام‌شده با مقیاس ارزیابی مبتنی بر ساختار باور، مقبولیت بالاتری نسبت به روش‌های معمول ارزیابی دارد.

جدول شمارهٔ یازده- خروجی آزمون فریدمن در بررسی مقبولیت نتایج مقیاس‌های مختلف ارزیابی در پرسش‌نامهٔ دوم

	مقیاس لیکرت	عبارات زبانی فازی	مقیاس مشابهت بصری	ساختار باور
تعداد نمونه	٤۵	٤۵	٤۵	٤۵
میانه	۱	۳	۳	٤
مجموع رتبه‌ها	۸٤	۱۱۷	۱۱۸	۱۳۱
میانگین رتبه‌ها	۱,۸۷	۲,٦	۲,٦۲	۲,۹۱
آماره کای دو	۱٦,۵۸			
درجه آزادی	۳			
p-value	۰,۰۰۰۰			

۵،۴ بررسی فرضیهٔ چهارم: مقایسهٔ سهولت اعلام نظر با مقیاس‌های ارزیابی

در بررسی سهولت اعلام نظر با مقیاس‌های ارزیابی، با توجه به تفاوت داده‌های به‌دست‌آمده در پرسش‌نامهٔ اول و دوم، نتایج به تفکیک مورد بررسی قرار گرفت. این نتایج در جدول ۱۲ و ۱۳ آمده است.

جدول شمارهٔ دوازده- نتایج آزمون فریدمن در بررسی سهولت کار با مقیاس‌های ارزیابی پرسش‌نامهٔ اول

	مقیاس لیکرت	عبارات زبانی فازی	مقیاس مشابهت بصری	ساختار باور
تعداد نمونه	۵	۵	۵	۵
میانه	۵	۵	۲	٤
مجموع رتبه‌ها	۱٦	۱۷،۵	۷	۹۰۵
میانگین رتبه‌ها	۳،۲	۳،۵	۱،٤	۱،۹
آماره کای دو	۱۰،٤۳			
درجه آزادی	۳			
p-value	۰،۰۱۵			

جدول شمارهٔ سیزده- نتایج آزمون فریدمن در بررسی سهولت کار با مقیاس‌های ارزیابی در پرسش‌نامهٔ دوم

	مقیاس لیکرت	عبارات زبانی فازی	مقیاس مشابهت بصری	ساختار باور
تعداد نمونه	۵	۵	۵	۵
میانه	٤	٤	۳	۵
مجموع رتبه‌ها	۱۱	۱۵	٦	۱۸
میانگین رتبه‌ها	۲،۲	۳	۱،۲	۳،٦
آماره کای دو	۱۲،۱۵			
درجه آزادی	۳			
p-value	۰،۰۰٦			

نتایج آزمون فریدمن نشان می‌دهد در سطح خطای ۵% در هر دو پرسش‌نامه، بین راحتی پاسخ‌گویی با استفاده از مقیاس‌های مختلف ارزیابی تفاوت معنادار وجود دارد. در پرسش‌نامهٔ اول که از سؤالات ساده و شفاف استفاده شده است، به ترتیب مقیاس عبارات فازی که به درک انسان نزدیک‌تر است و مقیاس لیکرت از نظر سهولت پاسخ‌گویی بیشترین رتبه را دارد. اما در پرسش‌نامهٔ دوم که سؤالات آن پیچیده‌تر است، اولویت ارزیاب‌ها، اعلام نظر در قالب توابع باور بوده است.

٦ جمع‌بندی و نتیجه‌گیری

در این پژوهش یک ابزار جدید ارزیابی و اعلام نظر در مدل‌های ارزیابی عملکرد روی مقیاس‌های ارزیابی مبتنی بر ساختار توابع باور ارائه شد که قابلیت دریافت نظرات توأم با عدم قطعیت ناشی از نقص دانش ارزیاب و همچنین قطعیت ناشی از ماهیت نادقیق پدیده مورد ارزیابی را دارا است. در تحلیل‌های انجام‌شده در این مقاله کارایی و اثربخشی این مقیاس و روش اعلام نظر بررسی شد.

نتایج به‌دست‌آمده از کاربرد عملی این روش ارزیابی در کنار سایر مقیاس‌های متداول دریافت نظر از ارزیاب‌ها نشان می‌دهد زمانی که نظرسنجی در مورد سؤالات، شفاف و بدون ابهام و پیچیدگی انجام می‌شود، هیچ‌گونه تفاوت معناداری در خروجی حاصل از مقیاس‌های مختلف ارزیابی وجود ندارد. اما در مورد سؤالات مبهم و مواردی که ارزیاب‌ها ممکن است اطلاعات کاملی در آن حوزه‌ها نداشته باشند، بین نظرات دریافت شده با استفاده از مقیاس‌های مختلف ارزیابی، تفاوت معنادار وجود دارد. برای تعیین روش مناسب اعلام نظر در این شرایط، بررسی مقبولیت نتایج به‌دست‌آمده انجام گرفت. در مقایسه مقبولیت نتایج به‌دست‌آمده از مقیاس‌های مختلف ارزیابی، مشخص شد که نتایج حاصل از نظرسنجی مبتنی بر ساختار باور مقبولیت بالاتری دارد.

از منظر راحتی کار با مقیاس‌های ارزیابی برای افراد اعلام‌نظرکننده، نتایج نشان می‌دهد ارزیاب‌ها زمانی که سؤالات شفاف از آن‌ها پرسیده می‌شود یا اطلاعات کافی برای پاسخ-گویی به سؤالات در اختیار دارند، ترجیح می‌دهند

از مقیاس‌های معمول ارزیابی مانند عبارات فازی و مقیاس لیکرت استفاده نمایند؛ اما در شرایطی که با سؤالات مبهم و پیچیده مواجه‌اند یا اطلاعات کافی جهت پاسخ‌گویی به سؤالات در اختیار ندارند، ارزیابی در ساختار توابع باور را ترجیح می‌دهند؛ لذا می‌توان در پرسش‌نامه‌ها امکان اعلام نظر معمول

در قالب مقیاس لیکرت و اعلام نظر در ساختار توابع باور را برای ارزیاب‌ها فراهم آورد تا اینکه در شرایطی که ارزیاب احساس می‌کند سؤالات از پیچیدگی برخوردار است یا اطلاعات آن‌ها جهت اعلام نظر قطعی، ناقص است از توابع باور جهت اعلام نظر استفاده کنند.

پیشنهادها و مطالعات آینده

نتایج نشان می‌دهد در ارزیابی‌هایی که با استفاده از مقیاس‌های متداول انجام می‌گیرد، به‌منظور دریافت نتایج صحیح از فرآیند ارزیابی، با ساده‌سازی و تک‌بعدی کردن سؤالات، می‌توان تا حد امکان به شفافیت سؤال و ساده شدن اعلام نظر ارزیاب کمک کرد. در غیر این‌صورت، زمانی که سؤالات و مفاهیم مورد ارزیابی پیچیده و توام با تغییرپذیری و عدم قطعیت است، مطابق بررسی‌های انجام‌شده در این پژوهش پیشنهاد می‌شود از ساختار توابع باور به‌عنوان یک گزینهٔ مناسب جهت استفاده در این شرایط بهره گرفته شود.

بر اساس یافته‌های پژوهش پیشنهاد می‌گردد در ارزیابی‌ها و نظرسنجی‌هایی که احتمال می‌رود پاسخ‌دهندگان اطلاعات کامل جهت اعلام نظر قطعی و مشخص نداشته باشند و همچنین تعداد پاسخ‌دهندگان در آن محدود است، جهت افزایش دقت و اعتبار نتایج، امکان اعلام نظر بر اساس ساختار توابع باور (به‌عنوان یک گزینهٔ اختیاری) در کنار روش‌های معمول ارزیابی فراهم گردد.

همچنین با توجه به بالابودن حجم محاسبات در تئوری شواهد، پیشنهاد می‌شود در مطالعات آتی، امکان کاربرد سایر قوانین ترکیب شواهد که در حوزهٔ تئوری شواهد ارائه شده‌اند، با هدف کاهش حجم محاسبات مرتبط با تجمیع شواهد بررسی گردد.

منابع

Azwir, H. H., & Kalinggo, B. A. (2019). Multistage Fuzzy Inference System for Solving Problems in Performance Appraisal. *2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, 200–205. https://doi.org/10.1109/ICSECC.2019.8907212

Baraldi, P., Mangili, F., & Zio, E. (2015). A belief function theory based approach to combining different representation of uncertainty in prognostics. *Information Sciences*, *303*, 134–149. https://doi.org/https://doi.org/10.1016/j.ins.2014.12.051

Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001a). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology*, *86*(5), 965.

Borman, W. C., Buck, D. E., Hanson, M. A., Motowidlo, S. J., Stark, S., & Drasgow, F. (2001b). An examination of the comparative reliability, validity, and accuracy of performance ratings made using computerized adaptive rating scales. *Journal of Applied Psychology*, *86*(5), 965–973.

AbuDahab, K., Xu, D. L., & Chen, Y. W. (2016). A new belief rule base knowledge representation scheme and inference methodology using the evidential reasoning rule for evidence combination. *Expert Systems with Applications*, *51*, 218–230. https://doi.org/10.1016/j.eswa.2015.12.013

Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behavior, personality, and attitudes*. John Wiley & Sons.

Arbaiy, N., & Suradi, Z. (2007). Staff performance appraisal using fuzzy evaluation. *Artificial Intelligence and Innovations 2007: From Theory to Applications*, 195–203. https://doi.org/10.1007/978-0-387-74161-1_21

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, *77*(6), 836–874. https://doi.org/10.1037/0021-9010.77.6.836

Awogbami, G., Agana, N., Nazmi, S., & Homaifar, A. (2018). A New Combination Rule Based on the Average Belief Function. *SoutheastCon 2018*, 1–5. https://doi.org/10.1109/SECON.2018.8478815

Brutus, S. S. (2010). Words versus numbers: A theoretical exploration of giving and receiving narrative comments in performance appraisal. *Human Resource Management Review*, *20*(2), 144–157. https://doi.org/10.1016/j.hrmr.2009.06.003

Chang, J.-R., Cheng, C.-H., & Chen, L.-S. (2007). A fuzzy-based military officer performance appraisal system. *Applied Soft Computing*, *7*(3), 936–945. https://doi.org/http://dx.doi.org/10.1016/j.asoc.2006.03.003

Cotton, D. R. E., Nash, T., & Kneale, P. (2017). Supporting the retention of non-traditional students in Higher Education using a resilience framework. *European Educational Research Journal*, *16*(1), 62–79.

De La Rosa De Sáa, S., Gil, M. Á., García, M. T. L., & Lubiano, M. A. (2013). Fuzzy rating vs. fuzzy conversion scales: An empirical comparison through the MSE. *Advances in Intelligent Systems and Computing*, *190 AISC*, 135–143. https://doi.org/10.1007/978-3-642-33042-1_15

Deng, Y., Sadiq, R., Jiang, W., & Tesfamariam, S. (2011). Risk analysis in a linguistic environment: A fuzzy evidential reasoning-based approach. *Expert Systems with Applications*, *38*(12), 15438–15446. https://doi.org/10.1016/j.eswa.2011.06.018

Dezert, J., Wang, P., & Tchamova, A. (2012). On the validity of Dempster-Shafer theory. *Information Fusion (FUSION), 2012 15th International Conference On*, 655–660.

Ding, S., Ma, X.-J., & Yang, S.-L. (2012). A software trustworthiness evaluation model using objective weight based evidential reasoning approach. *Knowledge and Information Systems*, *33*(1), 171–189.

Esen, H., Hatipoğlu, T., & Boyacı, A. İ. (2016). A Fuzzy Approach for Performance Appraisal: The Evaluation of a Purchasing Specialist. In *Studies in Computational Intelligence* (Vol. 620, pp. 235–250). Springer. https://doi.org/10.1007/978-3-319-26393-9_14

Espinilla, M., de Andrés, R., Martínez, F. J., & Martínez, L. (2013). A 360-degree performance appraisal model dealing with heterogeneous information and dependent criteria. *Information Sciences*, *222*, 459–471. https://doi.org/http://dx.doi.org/10.1016/j.ins.2012.08.015

Fu, C., Yang, J.-B., & Yang, S.-L. (2015). A group evidential reasoning approach based on expert reliability. *European Journal of Operational Research*, *246*(3), 886–893. https://doi.org/http://dx.doi.org/10.1016/j.ejor.2015.05.042

Fu, C., & Yang, S. (2012). An evidential reasoning based consensus model for multiple attribute group decision analysis problems with interval-valued group consensus requirements. *European Journal of Operational Research*, *223*(1), 167–176. https://doi.org/10.1016/j.ejor.2012.05.048

Golec, A., & Kahya, E. (2007). A fuzzy model for competency-based employee evaluation and selection. *Computers & Industrial Engineering*, *52*(1), 143–161.

Golman, R., & Bhatia, S. (2012). Performance evaluation inflation and compression. *Accounting, Organizations and Society*, *37*(8), 534–543.

Gürbüz, T., & Albayrak, Y. E. (2014). An engineering approach to human resources performance evaluation: Hybrid MCDM application with interactions. *Applied Soft Computing Journal*, *21*, 365–375. https://doi.org/10.1016/j.asoc.2014.03.025

Ha-Duong, M. (2008). Hierarchical fusion of expert opinions in the Transferable Belief Model, application to climate sensitivity. *International Journal of Approximate Reasoning*, *49*(3), 555–574. https://doi.org/10.1016/j.ijar.2008.05.003

Hedge, J. W., & Teachout, M. S. (2000). Exploring the Concept of Acceptability as a Criterion for Evaluating Performance Measures. *Group & Organization Management*, *25*(1), 22–44. https://doi.org/10.1177/1059601100251003

Hesketh, T., Pryor, R., & Hesketh, B. (1988). An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. *International Journal of Man-Machine Studies*, *29*(1), 21-35.

Hoffman, B. J., Gorman, C. A., Blair, C. A., Meriac, J. P., Overstreet, B., & Atchley, E. K. (2012). Evidence for the Effectiveness of an Alternative Multisource Performance Rating Methodology. *Personnel Psychology*, *65*(3), 531–563. https://doi.org/10.1111/j.1744-6570.2012.01252.x

Hooft, E. A. J., Flier, H., & Minne, M. R. (2006). Construct Validity of Multi- Source Performance Ratings: An Examination of the Relationship of Self- , Supervisor- , and Peer- Ratings with Cognitive and Personality Measures. *International Journal of Selection and Assessment*, *14*(1), 67–81.

Huang, L., Ruan, S., & Denœux, T. (2021). Belief Function-Based Semi-Supervised Learning For Brain Tumor Segmentation. *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 160–164. https://doi.org/10.1109/ISBI48211.2021.9433885

Jamsandekar, S. S., & Mudholkar, R. R. (2013). *Performance Evaluation by Fuzzy Inference Technique*. *2*, 158–164.

Jelley, R. B., Goffm, R. D., & Goffin, R. D. (2001). Can performance-feedback accuracy be improved? Effects of rater priming and rating-scale format on rating accuracy. *Journal of Applied Psychology*, *86*(1), 134.

Jiang, W., Wei, B., Xie, C., & Zhou, D. (2016). An evidential sensor fusion method in fault diagnosis. *Advances in Mechanical Engineering*, *8*(3), 1-7.

Jin, L., Fang, X., & Xu, Y. (2015). A Method for Multi-attribute Decision Making Under Uncertainty Using Evidential Reasoning and Prospect Theory. *International Journal of Computational Intelligence Systems*, *8*(sup1), 48–62. https://doi.org/10.1080/18756891.2015.1129578

John Bernardin, H., Thomason, S., Ronald Buckley, M., & Kane, J. S. (2016). Rater rating- level bias and accuracy in performance appraisals: The impact of rater personality, performance management competence, and rater accountability. *Human Resource Management*, *55*(2), 321-340.

Klein, J., & Colot, O. (2011). Singular sources mining using evidential conflict analysis. *International Journal of Approximate Reasoning*, *52*(9), 1433–1451. https://doi.org/10.1016/j.ijar.2011.08.005

Lefevre, E., Colot, O., & Vannoorenberghe, P. (2002). Belief functions combination and conflict management. *Information Fusion Journal*, *3*(2), 149–162.

Lefèvre, E., & Elouedi, Z. (2013). How to preserve the conflict as an alarm in the combination of belief functions? *Decision Support Systems*, *56*(1), 326–333. https://doi.org/10.1016/j.dss.2013.06.012

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*(6), 881–905.

Liu, Z. G., Dezert, J., Pan, Q., & Mercier, G. (2011). Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. *Decision Support Systems*, *52*(1), 133–141. https://doi.org/10.1016/j.dss.2011.06.002

Lubiano, M. A., de La Rosa De Sáa, S., Montenegro, M., Sinova, B., & Gil, M. Á. (2016a). Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale? *Information Sciences*, *360*, 131–148. https://doi.org/10.1016/j.ins.2016.04.029

Lubiano, M. A., Montenegro, M., Sinova, B., de la Rosa de Sáa, S., & Gil, M. Á. (2016b). Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. *European Journal of Operational Research*, *251*(3), 918–929. https://doi.org/http://dx.doi.org/10.1016/j.ejor.2015.11.016

Macwan, N., & Sajja, D. P. S. (2013). Performance appraisal using fuzzy evaluation methodology. *International Journal of Engineering and Innovative Technology*, *3*(3), 324-329.

Manoharan, T. R., Muralidharan, C., & Deshmukh, S. G. (2011). An integrated fuzzy multi-attribute decision-making model for employees' performance appraisal. *The International Journal of Human Resource Management*, *22*(3), 722–745. https://doi.org/10.1080/09585192.2011.543763

Maseleno, A., Hasan, Md. M., Tuah, N., & Tabbu, C. R. (2015). Fuzzy Logic and Mathematical Theory of Evidence to Detect the Risk of Disease Spreading of Highly Pathogenic Avian Influenza H5N1. *Procedia Computer Science*, *57*, 348–357. https://doi.org/http://dx.doi.org/10.1016/j.procs.2015.07.349

Meenakshi, G. (2012). Multi source feedback based performance appraisal system using Fuzzy logic decision support system. *International Journal on Soft Computing*, *3*(1), 91–106. https://doi.org/10.5121/ijsc.2012.3108

Min, N., Sirisrisakulchai, J., Sriboonchitta, S. (2015). Forecasting Tourist Arrivals to Thailand Using Belief Functions. In: Huynh, VN., Kreinovich, V., Sriboonchitta, S., Suriya, K. (eds) *Econometrics of Risk. Studies in Computational Intelligence*, vol 583. Springer, Cham. https://doi.org/10.1007/978-3-319-13449-9_24

Moon, C., Lee, J., & Lim, S. (2010). A performance appraisal and promotion ranking system based on fuzzy logic: An implementation case in military organizations. *Applied Soft Computing*, *10*(2), 512–519. https://doi.org/http://dx.doi.org/10.1016/j.asoc.2009.08.035

Murphy, C. K. (2000). Combining belief functions when evidence conflicts. *Decision Support Systems*, *29*(1), 1–9. https://doi.org/10.1016/S0167-9236(99)00084-6

Murphy, K. R. (2008a). Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance. *Industrial and Organizational Psychology*, *1*(2), 148–160. https://doi.org/10.1111/j.1754-9434.2008.00030.x

Murphy, K. R. (2008b). Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance. *Industrial and Organizational Psychology*, *1*(2), 148–160. https://doi.org/10.1111/j.1754-9434.2008.00030.x

Murphy, K. R., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Sage.

Neogi, A., Mondal, A. C. A. A. C. A., & Mandal, S. S. K. (2011). A cascaded fuzzy inference system for university non-teaching staff performance appraisal. *Journal of Information Processing Systems*, *7*(4), 595–612. https://doi.org/10.3745/JIPS.2011.7.4.595

Ng, C. Y. (2016). Evidential reasoning-based Fuzzy AHP approach for the evaluation of design alternatives' environmental performances. *Applied Soft Computing Journal*, *46*, 381–397. https://doi.org/10.1016/j.asoc.2016.05.022

Nowack, K. M., & Mashihi, S. (2012). Evidence-based answers to 15 questions about leveraging 360-degree feedback. *Consulting Psychology Journal: Practice and Research*, *64*(3), 157-182.

Oberkampf, W. L., DeLand, S. M., Rutherford, B. M., Diegert, K. v., & Alvin, K. F. (2002). Error and uncertainty in modeling and simulation. *Reliability Engineering & System Safety*, *75*(3), 333–357. https://doi.org/10.1016/S0951-8320(01)00120-X

Oukhellou, L., Debiolles, A., Denœux, T., & Aknin, P. (2010). Fault diagnosis in railway track circuits using Dempster–Shafer classifier fusion. *Engineering Applications of Artificial Intelligence*, *23*(1), 117–128. https://doi.org/http://dx.doi.org/10.1016/j.engappai.2009.06.005

Paksoy, A., & Göktürk, M. (2011). Information fusion with dempster-shafer evidence theory for software defect prediction. *Procedia Computer Science*, *3*, 600–605.

Qiang, C., & Deng, Y. (2022a). A new correlation coefficient of mass function in evidence theory and its application in fault diagnosis. *Applied Intelligence*, *52*(7), 7832–7842. https://doi.org/10.1007/s10489-021-02797-2

Qiang, C., & Deng, Y. (2022b). A new correlation coefficient of mass function in evidence theory and its application in fault diagnosis. *Applied Intelligence*, *52*, 1–11. https://doi.org/10.1007/s10489-021-02797-2

Saris, W. E., & van Meurs, A. (1990). Evaluation of measurement instruments by meta-analysis of multitrait multi-method studies. *Proceedings Amsterdam: North Holland*.

Schubert, J. (2011). Conflict management in Dempster-Shafer theory using the degree of falsity. *International Journal of Approximate Reasoning*, *52*(3), 449–460. https://doi.org/10.1016/j.ijar.2010.10.004

Sentz, K., & Ferson, S. (2002). *Combination of evidence in Dempster-Shafer theory* (Vol. 4015). Citeseer, Sandia National Laboratories.

Shaout, A., Yousif, M. K., & Khalid Yousif, M. (2014). Employee Performance Appraisal System Using Fuzzy Logic. *International Journal of Computer Science and Information Technology*, *6*(4), 1–19. https://doi.org/10.5121/ijcsit.2014.6401

Smets, P. (2005). Decision making in the TBM: The necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, *38*(2), 133–147. https://doi.org/10.1016/j.ijar.2004.05.003

Smets, P. (2007). Analyzing the combination of conflicting belief functions. *Information Fusion*, *8*(4), 387–412. https://doi.org/10.1016/j.inffus.2006.04.003

Tang, D., Wong, T. C., Chin, K. S., & Kwong, C. K. (2014). Evaluation of user satisfaction using evidential reasoning-based

methodology. *Neurocomputing*, *142*, 86–94. https://doi.org/http://dx.doi.org/10.1016/j.neucom.2014.01.055

Taroun, A., & Yang, J.-B. (2011). Dempster-Shafer Theory of Evidence: Potential usage for decision making and risk analysis in construction project management. *The Built & Human Environment Review*, 4(1), 155-166.

Tong, Z., Xu, P., & Denœux, T. (2021). An evidential classifier based on Dempster-Shafer theory and deep learning. *Neurocomputing*, *450*, 275–293. https://doi.org/https://doi.org/10.1016/j.neucom.2021.03.066

Turgut, H., & Mert, I. S. (2014). Evaluation of Performance Appraisal Methods through Appraisal Errors by Using Fuzzy VIKOR Method. *International Business Research*, *7*(10), 170–178. https://doi.org/10.5539/ibr.v7n10p170

Vasu, J. Z., Deb, A. K., & Mukhopadhyay, S. (2015). MVEM-based fault diagnosis of automotive engines using Dempster–Shafer theory and multiple hypotheses testing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(7), 977-989.

Wagner, S. H., & Goffin, R. D. (1997). Differences in accuracy of absolute and comparative performance appraisal methods. *Organizational Behavior & Human Decision Processes*, *70*(2), 95–103. https://doi.org/10.1006/obhd.1997.2698

Wang, Y., Zhang, K., & Deng, Y. (2019). Base belief function: an efficient method of conflict management. *Journal of Ambient Intelligence and Humanized Computing*, *10*(9), 3427–3437. https://doi.org/10.1007/s12652-018-1099-2

Wilson, N. (1993). Decision-making with Belief Functions and pignistic probabilities. In M. Clarke, R. Kruse, & S. Moral (Eds.), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty: European Conference ECSQARU '93 Granada, Spain, November 8--10, 1993 Proceedings* (pp. 364–371). Springer Berlin Heidelberg. https://doi.org/10.1007/BFb0028222

Woehr, D. J. (2008). On the Relationship Between Job Performance and Ratings of Job Performance: What Do We Really Know? *Industrial and Organizational Psychology*, *1*(2), 161–166.

Woehr, D. J., & Huffcutt, A. I. (1994a). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*(3), 189–205.

Woehr, D. J., & Huffcutt, A. I. (1994b). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*(3), 189–205.

Xiao, Z., Yang, X., Pang, Y., & Dang, X. (2012). The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster–Shafer evidence theory. *Knowledge-Based Systems*, *26*, 196–206. https://doi.org/http://dx.doi.org/10.1016/j.knosys.2011.08.001

Xu, D.-L. (2012). An introduction and survey of the evidential reasoning approach for multiple criteria decision analysis. *Ann Oper Res*, *195*(1), 163–187. https://doi.org/10.1007/s10479-011-0945-9

Yang, J. B., & Singh, M. G. (1994). An evidential reasoning approach for multiple-attribute decision making with uncertainty. *IEEE Transactions on systems, Man, and Cybernetics*, *24*(1), 1-18.

Ying-Feng, K., & Ling-Show, C. (2002). Using the fuzzy synthetic decision approach to assess the performance of university teachers in Taiwan. *International Journal of …*, *19*(4), 593–604.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to

approximate reasoning—I. *Information Sciences*, *8*(3), 199–249.

Zadeh, L. A. (1986). A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, *7*(2), 85.