

Research paper/Оригинальная статья

<https://doi.org/10.51176/1997-9967-2022-3-108-123>

МРНТИ 06.61.33

JEL: I28, I38, R58



A Critical Analysis of Regional Program Evaluation Practice in Kazakhstan

Nazym N. Battalov^{1*}

¹ *Al-Farabi Kazakh National University, 71 al-Farabi Ave., 050040, Almaty, Kazakhstan*

Abstract

In recent years the development of public program evaluation has received growing attention in Kazakhstan. An institutional and legal base for program evaluation has been established. However, the examination of literature has evidenced that there have been rudimentary attempts to interrogate evaluation practice, particularly at the regional level. It is still not well known how effective or valuable it is. It is imperative to run a diagnostic and assess the evaluation system to answer this question. This article aims to evaluate the quality of regional program evaluation practice in Kazakhstan. It applies a meta-evaluation tool to understand the extent to which such practice complies with three fundamental and recognized evaluation standards: namely, value, validity, and utility. As a sample, the study used evaluation reports conducted by regional Audit commissions. This research is the first attempt to apply established evaluation standards to the Kazakhstani context. Therefore, it was assumed that some discrepancies with the standards may occur. Having confirmed this hypothesis, the findings indicate that regional program evaluation falls far short of these standards. The paper identified many conceptual and methodological problems, which seriously compromise the validity and soundness of evaluation practice. It is expected that it will stimulate discussion in academic and subject matter expert circles. Furthermore, having identified key areas for improvement, the study may help reform the evaluation field and contribute to better policy- and decision-making, thus saving taxpayers' money and improving people's wellbeing. In the end, the research put forward several recommendations for strengthening evaluation practice.

Keywords: Evaluation, Effectiveness, Public Programs, Kazakhstan, Meta-Evaluation, Practice

For citation: Battalov, N.N. (2022). A Critical Analysis of Regional Program Evaluation Practice in Kazakhstan. *Economics: the Strategy and Practice*, 17(3), 108-123, <https://doi.org/10.51176/1997-9967-2022-3-108-123>

* **Corresponding author:** Battalov N.N. – PhD candidate, Al-Farabi Kazakh National University, 71, al-Farabi Ave., 050040, Almaty, Kazakhstan, 87023706934, e-mail: nnbattalov@gmail.com

Conflict of interests: the author declares that there is no conflict of interest.

Financial support: The study was not sponsored (own resources).

The article received: 30.05.2022

The article approved for publication: 30.06.2022

Date of publication: 30.09.2022

Қазақстандағы өңірлік бағдарламаларды бағалау практикасына сыни талдау

Батталов Н.Н.^{1*}

¹ *әл-Фараби Қазақ ұлттық университеті, әл-Фараби 71, 050040, Алматы, Қазақстан*

Түйін

Соңғы жылдары Қазақстанда мемлекеттік бағдарламаларды бағалау тақырыбына назар артуда. Бағдарламаларды бағалаудың институционалдық және құқықтық базасы қалыптастырылды. Алайда, әдебиеттерге шолу бағалау практикасын зерттеуге, әсіресе аймақтық деңгейде, аз ғана талпыныс жасалғанын көрсетеді. Оның қаншалықты тиімді және пайдалы екендігі аз зерттелген. Бұл сұраққа жауап беру үшін бағалау жүйесін диагностикалау және бағалау өте маңызды. Осы мақаланың мақсаты - Қазақстандағы аймақтық бағдарламаларды бағалау тәжірибесінің сапасын бағалау. Мақалада қолданыстағы бағалау тәжірибесінің үш негізгі және танылған стандартқа, атап айтқанда құндылықтық аспекті, жарамдылық және пайдалылық стандарттарына қаншалықты сәйкес келетінін түсіну үшін мета-бағалау құралы қолданылады. Іріктеме ретінде өңірлік Тексеру комиссияларының дайындаған бағалау туралы есептер пайдаланылды. Аталған зерттеу қазақстандық мәнмәтінде бағалаудың танылған стандарттары мен мета-бағалау стратегияларын қолданудың алғашқы әрекеті болып табылады. Осыған байланысты стандарттардан кейбір айырмашылықтар болуы мүмкін деп болжанған болатын. Алынған нәтижелер, осы гипотезаны растай отырып, бағдарламаларды аймақтық деңгейде бағалау осы стандарттардан едәуір артта қалатынын көрсетті. Мақалада аймақтық бағдарламаларды бағалаудың дұрыстығы мен негізділігіне нұқсан келтіретін көптеген тұжырымдамалық және әдіснамалық мәселелер анықталды. Мақала академиялық және сараптамалық топтардағы пікірталастарды ынталандырады деп күтілуде. Сонымен қатар, жетілдіруді қажет ететін негізгі бағыттарды анықтай отырып, мақала бағалау саласын реформалауға және саясатты және шешімдерді сапалы әзірлеуге ықпал етуі мүмкін. Осылайша салық төлеушілердің ақшасын үнемдеуге және адамдардың әл-ауқатын жақсартуға әсерін тигізе алады. Соңында мақалада бағалау практикасын жаңғырту бойынша бірқатар ұсынымдар беріледі.

Түйін сөздер: бағалау, тиімділік, мемлекеттік бағдарламалар, Қазақстан, метабағалау, практика

Дәйексөз үшін: Батталов Н.Н. (2022). Қазақстандағы өңірлік бағдарламаларды бағалау практикасына сыни талдау. Экономика: стратегия және практика, 17(3), 108-123, <https://doi.org/10.51176/1997-9967-2022-3-108-123>

* **Хат-хабаршы авторы:** Батталов Н.Н. – PhD докторанты, әл-Фараби Қазақ ұлттық университеті, әл-Фараби 71, 050040, Алматы, Қазақстан, 87023706934, e-mail: nnbattalov@gmail.com

Мүдделер қақтығысы: автор мүдделер қақтығысының жоқтығын мәлімдейді.

Қаржыландыру: Зерттеуге демеушілік қолдау көрсетілмеді (меншікті ресурстар).

Мақала редакцияға түсті: 30.05.2022

Жариялау туралы шешім қабылданды: 30.06.2022

Жарияланды: 30.09.2022

Критический анализ практики оценки региональных программ в Казахстане

Батталов Н.Н.^{1*}

¹ *Казахский национальный университет им. аль-Фараби, пр. аль-Фараби 71, 050040,
Алматы, Казахстан*

Аннотация

За последние годы тема оценки государственных программ получает все большее внимание в Казахстане. Была сформирована институциональная и правовая база оценки программ. Однако, обзор литературы свидетельствует о том, что были предприняты лишь малые попытки исследования практики оценки, в особенности на региональном уровне. Остается малоизученным то, насколько она эффективна и полезна. Для ответа на этот вопрос, безусловно, важно диагностировать и оценить саму систему оценки. Цель данной статьи – оценить качество практики оценки региональных программ в Казахстане. В статье использован инструмент метаоценки для понимания того, насколько действующая практика соответствует трем фундаментальным и признанным стандартам оценки: ценностное измерение, валидность и полезность. В качестве выборки использованы отчеты об оценке, составленные региональными Ревизионными комиссиями. Данное исследование является первой попыткой применения признанных стандартов оценки и стратегий метаоценки в казахстанском контексте. Ввиду этого, предполагалось, что могут иметь место некие расхождения со стандартами. Полученные результаты, подтвердив данную гипотезу, свидетельствуют о том, что оценка программ на региональном уровне значительно отстает от данных стандартов. Статья определила множество концептуальных и методологических проблем, которые серьезно подрывают валидность и обоснованность оценки региональных программ. Ожидается, что статья будет стимулировать дискуссии в академических и экспертных кругах. Более того, определив ключевые направления, требующие совершенствования, статья может помочь в реформировании сферы оценки и способствовать более качественной разработке политик и решений, тем самым экономя деньги налогоплательщиков и улучшая благополучие людей. В завершении статья выдвигает ряд рекомендаций по модернизации практики оценки.

Ключевые слова: оценка, эффективность, государственные программы, Казахстан, метаоценка, практика

Для цитирования: Батталов Н.Н. (2022). Критический анализ практики оценки региональных программ в Казахстане. Экономика: стратегия и практика, 17(3), 108-123, <https://doi.org/10.51176/1997-9967-2022-3-108-123>

* **Корреспондирующий автор:** Батталов Н.Н. – PhD докторант, Казахский национальный университет им. аль-Фараби, пр. аль-Фараби 71, 050040, Алматы, Казахстан, 87023706934, e-mail: nnbattalov@gmail.com

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Финансирование: Исследование не имело спонсорской поддержки (собственные ресурсы).

Статья поступила в редакцию: 30.05.2022

Принято решение о публикации: 30.06.2022

Опубликовано: 30.09.2022

Introduction

Program evaluation in Kazakhstan has become an essential aspect of public management (Nygmetov, 2014). Early in 2020, under the Office of the President of Kazakhstan, the Centre for Analysis and Monitoring was created and assessed the Government's programs and reforms. An analogous structure, named the Centre for Evaluation of Public Programs and Reforms, was founded within the ruling political party Amanat (Nur Otan until March 2022) in 2019. The establishment of such institutions, when there are already evaluation bodies operating at the national (Accounts committee) and regional level (Audit commissions), can be viewed as an indirect indication of the insufficient effectiveness in the existing program evaluation practice.

Furthermore, both mentioned Centers have tended to focus on state-level interventions. The same discourse can be observed in expert and academic circles when discussions center around nationally implemented programs (Vechkinzova, 2008; Bopiyeva & Kazakova, 2009a; Bopiyeva & Kazakova, 2009b; Gayfutdinova, 2017; Pokidayev, 2020; Kaldiyarov & Turgambekova, 2019; Yessimova, 2009; Baimbetov, 2019). However, public spending on programs at regional or local (such terms are used interchangeably) levels represents a considerable portion of the country's total national budget. Therefore, investigating the practice of evaluating regional programs deserves considerably more attention, not least for financial accountability purposes (OECD, 2021).

Some concerns were expressed by the President of Kazakhstan Kassym-Jomart Tokayev, who said "The Government develops reforms, implements them, and then evaluates the quality itself. This situation needs to be changed" (Akorda, 2020, p.1). Furthermore, Mr. Tokayev recently called for developing a methodology for assessing state expenditures' social and economic effectiveness, thus emphasizing the importance of improving evaluation in the public sector.

It should be acknowledged that evaluation is a relatively new field in Kazakhstan's public policy arena, with significant developments in this domain has taken place during the last decade. As with any professional enterprise, it should be subjected to proper scrutiny to identify if practices suffer from flaws or mistakes. Building on the theoretical and methodological literature, this research will attempt to fill the existing gap in the literature and examine the practice of

evaluation of regional programs¹.

Thus, this paper aims to critically analyze and identify areas for improvement in the program evaluation at the local level. The author will utilize a meta-evaluation checklist to address it and assess three fundamental assessment standards: values, validity, and utility.

To achieve this purpose, the following research questions have been put forward:

1. How well does regional program evaluation conform with established evaluation standards?
2. How justified and appropriate are the values used in program evaluation?
3. How valid are program evaluation design and conclusions?
4. How useful are program evaluation conclusions and recommendations?

Literature review

The literature examination illustrates a shortage of systematic and in-depth studies devoted to the program evaluation, especially at the regional level. One of the few attempts to question evaluation practice in Kazakhstan was made by Nygmetov (2014), who argues that, although many efforts have been made to establish an evaluation system, their effectiveness is far from satisfactory since evaluation is perceived as a form of control and monitoring and not aimed at assessing a real impact of a program. Kari (2015) believes that the potential of evaluation remains unrealized and notes that there are inconsistencies between existing methodologies and the overlapping functions of evaluation bodies that prevent evaluation from being conducted systematically. Kari (2015) concurs with Nygmetov (2014) and points out that current evaluation practice emphasizes assessing short-term outcomes, while impact evaluation is not well developed.

Studies of specific programs usefully illuminate the limitations of evaluation practice. Pritvorova and Bektleyeva (2017) investigated 'Youth Internship' programs providing new graduates with six-month paid internships in state-owned organizations. They argue that while the program has been evaluated based on the number of participants, its longer-term effects (i.e., the job prospects of participants) were not considered.

¹ According to the latest amendments in the system of state planning in Kazakhstan in 2021, state programs have been renamed national projects. However, for the purposes of this paper, the term "program" will be used since the sample consisted of state programs, which were in force before the adoption of national projects. The term "regional program" refers here to any state-funded program implemented at the regional level.

Similarly, dosekova et al. (2018) show that the evaluation of startup commercialization programs needs to be focused not only on input additionality (i.e., the resources spent by firms in addition to state subsidies) but also on outcome additionality. Thus, it is seen that the existing evaluation practice might not adequately address the complexity and multiple aspects of programs.

Some discussions have taken place about tailoring evaluation to the context of evaluated programs (Kaldiyarov & Turgambekova, 2019). In healthcare, for instance, Murzaliyeva and Karshalova (2018) argue that medical organizations in Kazakhstan, such as sanatoriums, in-patient facilities, and polyclinics, have different scopes and objectives; therefore, applying 'one-size-fits-all' indicators to assess their programs is not a justifiable way to judge their effectiveness express the importance of reforming evaluation approaches (Rakhmatullayeva et al., 2015). They discuss an alternative method for assessing the social impact of direct foreign investment in Kazakhstan based on the mathematical modeling method. Although they address state-level interventions, these studies raise important questions about the flexibility and adaptability of evaluation to various settings, including the regional level. To investigate this issue in more depth, it is vital to explore how evaluations are designed and the values that underpin them.

In addition to the above, there is also quite substantial literature written by OECD, which constitutes a valuable source of policy advice but says relatively little about evaluation practice. Nevertheless, its recommendations include creating an evaluation research unit within local executive bodies. It also indicates a weak culture of evaluation and continuous improvement in the public sector (OECD, 2021).

Audit commissions play a central role in program evaluation at the regional level - state entities mandated to conduct both audits and assessments. Therefore, it is vital to review the literature devoted to their activities. It has been found that few works deal with the limitations of the audit system regarding the conduct of program evaluation. The literature mainly investigates its potential role. It is agreed that its capacity has been substantially enhanced, but little is known about its effectiveness in practice. In this respect, some researchers (Dosayeva, 2019; Shakirova et al., 2019; Alibekova et al., 2019) state that the audit concept is a fundamentally new area for Kazakh science, and its capacity and weaknesses are yet to be explored. This illustrates the need for empirical and in-depth studies on the program evaluation dimension of the audit system.

Overall, the papers discussed provide some insights on evaluation practices in Kazakhstan, but they do not give an in-depth analysis of the problems identified, and more research is necessary. This is true for evaluating national and regional programs - although the former has received more attention in the literature. Nevertheless, examining both levels would not be feasible within a single article; thus, its scope is limited to the local level. Furthermore, since 2015, there have been changes in evaluation methodology and legislature (Adilet, 2020). Consequently, there is a clear need to reexamine the theory and practice of evaluation in light of these developments.

Methodology

The sample was drawn from publicly available evaluation reports conducted by regional Audit Commissions. It included three forms of evaluations: expert opinions, evaluation reports on budget implementation, and performance audit reports with evaluation sections. In performance audit reports, only the evaluation sections were subjected to the analysis.

Evaluation reports are published and openly accessible on the official websites of Audit Commissions. Initial searches demonstrated that some websites contained outdated reports, while several of them were not accessible. Therefore, missing reports were requested from relevant Commissions by completing an online form in the electronic government of Kazakhstan portal.

Initial data collection resulted in 87 evaluation reports. The reports represented 16 regions and cities of national significance. To identify which evaluation reports were applicable to answer the research question, the author has applied inclusion criteria, similar to Scott-Little et al.'s research (Scott-Little, 2002). Table 1 presents inclusion criteria and their description.

Table 1 - Inclusion criteria

Criteria	Description
Temporal relevancy	Evaluation reports must be published no earlier than 2018
Evaluation report characteristics	Evaluation reports must contain a detailed description of: (i) criteria and standards, their application; (ii) sources of data, methods of analysis (ii) conclusions; (iii) recommendations

Note - Adapted by the author from Scott-Little et al. (2002)

Legislature in the field of evaluation has gone through some amendments; therefore, the author considered only recent reports published after 2018 to reflect those changes. Secondly, reports had to be sufficiently detailed to allow for analysis. After applying these criteria, 39 reports were selected, which formed a sample for the meta-evaluation. Evaluation methodologies in Kazakhstan have not changed significantly since 2020. Therefore, the results of the paper may be relevant to this day.

Research methods and instrumentation

A systematic review of the evaluation reports was conducted to evaluate their quality by determining their adherence to the evaluation standards described below. As an instrument, the study has used the adapted and synchronized version of Scriven's Meta-evaluation checklist (Scriven, 1991) and Davidson's meta-evaluation tool (1995). Scriven's Meta-evaluation checklist (Scriven, 1995) includes five main criteria of quality: validity, utility, propriety, credibility, and cost. Assessing appropriateness, credibility, and cost standards were not feasible since this information was not reflected in evaluation reports.

Thus, two standards remained: validity and utility. Validity consists of multiple aspects, of which one is the values or criteria upon which the quality of the program is measured. Fournier noted that "criteria can make or break an evaluation because they...directly affect the validity of claims"(Fournier, 1995, p.19). Considering the significant contribution of values, it has been decided to examine this dimension separately. Consequently, the resulting checklist consisted of three standards: values, validity, and utility. Each of the standards was strengthened by adding relevant points from the Key Evaluation Checklist (Scriven, 1991) and the Program Evaluation Standards (JCSEE, 2011), particularly accuracy, utility, and evaluation accountability.

Results

Before presenting the findings, it is important to look at how evaluation is interpreted in the context of Kazakhstan. According to the Government Decree on the system of state planning (Adilet, 2020, p. 13), evaluation is "an instrument of determining the extent to which state programs achieve effectiveness and efficiency". Efficiency is understood as the accomplishment of best outputs and outcomes using the approved budget, while effectiveness implies the achievement of performance indicators prescribed by plans, programs, and strategies

(Adilet, 2020). This interpretation is clearly distinct from the widely recognized definition of evaluation, i.e., systematic determination of merit and worth of a thing (Scriven, 1991). The implications of this contrast are illustrated throughout the research.

Values

Where do values come from?

The data analysis has shown that evaluations of regional programs have drawn upon a minimal set of values, such as program targets, procedural requirements, and institutional and legislative norms. As evidence of failure to address multiple relevant values in evaluations, attention may be drawn to the business development program in Karaganda region, which was evaluated based upon the achievement of program targets, such as the increased number of recipients of entrepreneurship training and microloans for starting businesses (unpublished analytical papers of state bodies). However, the evaluation did not address the values from the perspective of potential impacts; specifically, it might be useful to look at how the program helped to enhance employment opportunities and overcome social and economic problems in the region.

The problem with limiting evaluation scope to pre-determined criteria is that evaluations may overlook numerous symptoms and causes contributing to the achievement of program objectives. As an illustration, the Healthcare Development Program of Karaganda region addressed only five targets (unpublished analytical papers of state bodies). However, the list could be extended to include other relevant objectives. For instance, Aymagambetov and Tyngisheva (2019) claim that the region has serious health issues associated with respiratory and circulatory systems.

Table 2 describes the criteria used to evaluate the Healthcare development program in Karaganda region.

It indicated an array of causes of cardiovascular diseases. However, the scope of evaluation was limited to assessing public education activities, which can be the solution for only one of the causes - specifically, a deficit of awareness about factors leading to cardiovascular diseases. It is important to note that the existing methodologies do not limit evaluators in selecting criteria. Evaluators can develop additional measures to assess programs using various sources (Adilet, 2020). However, the analysis has demonstrated that the potential of this practice has not been fully realized since evaluations have included in their repertoire only those

criteria already prescribed by methodologies (for example, program goals and legislative norms).

Table 2 - Criteria used in the evaluation of the Healthcare development program of Karaganda region

Criteria	Planned values (for 2018)	Actual values 2018
Life expectancy	72.6	71.7
Maternal mortality per 100 000	11.3	26.2
Child mortality per 1000	9.4	7.4
Mortality from malignant tumours per 100 000	93.1	93
Prevalence of HIV/AIDS in 15-49 age category	0.48%	0.434%

Note - Compiled by the author based on data (unpublished analytical papers of state bodies)

To summarize, the analysis revealed that there had been no evidence of (i) conducting a needs assessment, i.e., identifying and analyzing the priority needs of program impacts, or (ii) scrutinizing causes of problem areas of programs. This produces risks to the validity and accuracy of evaluation findings.

Achievement of program goals

Program objectives have acted as a primary criterion for determining the effectiveness of programs. Essentially, evaluations examined whether indicators were achieved and then calculated the percentage of achieving targets, which served as the basis for further conclusions. It is important to note that indicators are not differentiated or ranked. Such an objective-based approach can have serious problems since some objectives may be more significant or relevant than others; giving them equal weight may distort the validity of findings (Davidson, 2005; Stufflebeam & Coryn, 2014).

The above can be demonstrated in the following example. The Education Program of Karaganda region (unpublished analytical papers of state bodies) includes two different criteria upon which its effectiveness is measured (Table 3).

Both targets are relevant, however, the former does not reflect qualitative changes and illustrates only the program's outputs. The first target is more difficult to achieve than the second and more significant since PISA has proved to be an effective and valid knowledge assessment tool internationally (OECD, 2021). Failing to meet

the first target and achieve the second does not necessarily mean that the program performed poorly. However, the analysis has illustrated that evaluations did not grade targets depending on their significance, difficulty, or relevance.

Table 3 - Criteria of the Education program in Karaganda region

No.	Criteria of the education program in Karaganda Region
1	Achieving better results in the test of the OECD's Program for International Student Assessment (PISA) in the following subjects: Math, Reading, and Science
2.	Ensuring a rise in the percentage of students of vocational education institutions covered by dual education

Note - Compiled by the author based on data (OECD, 2021)

Another critique concerns the justifiability of criteria. An example could be the criterion of 'life expectancy' found in the evaluation of the healthcare program of West Kazakhstan region (unpublished analytical papers of state bodies). The program is unlikely to have significantly impacted it within a reporting period, as the target is global and influenced by various factors (Ho & Hendi, 2018). It must be assessed comprehensively and from a more long-term perspective. Therefore, applying this criterion to gauge program performance annually is questionable.

The research also found that even when objectives are in place, performing evaluation has not always been possible. Evaluations have primarily relied upon official statistical data to assess the achievement of goals and make claims about a program's effectiveness. When statistical data was unavailable, programs were not subjected to further investigation. This has been the case in many evaluations, which had the caveat that assessment of certain aspects of a program was not feasible due to the absence of official data (unpublished analytical papers of state bodies). Similarly, some programs' lack of measurable indicators has prevented evaluators from assessing them (unpublished analytical documents of state bodies). For instance, the program for controlling stray dogs and preventing zoonotic diseases in the Terekti district in West Kazakhstan region lacked any indicators and, therefore, no evaluative activities were undertaken.

To conclude, there was no evidence that various programs' goals were checked for relevance and significance. Further, dependence on program goals has seriously impaired the flexibility of evaluators.

Legislative norms and standards

The analysis of the data has demonstrated that an inordinate emphasis has been placed on the assessment of the adherence to legislative guidelines. This indicates the dilution of evaluation practice with elements of a compliance audit. For example, the expert report on evaluating the microloan and entrepreneurship development program in Mangistau region (unpublished analytical papers of state bodies) was predominantly assessed for conformity of program outputs with program specifications and lending regulations. Evaluators have examined the legality of granting microloans within the program by checking the eligibility of program participants. They then looked at whether grants complied with program conditions in creating new jobs by utilizing the funds received for the intended purpose. However, no inferences have been made regarding the impact and value of the program for the sphere of entrepreneurship and business climate in the region in general. The same trend can be seen in many evaluations, which, apart from assessing program objectives, verified the compliance of programs with provisions of the Budget Code and procedural norms for program planning and implementation (unpublished analytical papers of state bodies).

It is argued that this approach may only help to determine a program's merit or intrinsic value. To illustrate this, attention may be drawn to the evaluation of the innovation development program in Karaganda region (unpublished analytical papers of state bodies). It described some activities of the program, such as introducing an electronic ticket system on public transport and installing air pollution control sensors. The program may have conformed with its technical specifications and served its intended purpose; however, the evaluation did not investigate how the program activities had contributed to meeting the needs of the consumer population. The point is that even if legal requirements, technical specifications, or accepted standards of quality are followed, a program nonetheless 'might not be worthy' (Stufflebeam & Coryn, 2014, p. 9).

Validity***Evaluation logic***

The author investigated the basic logic underpinning evaluative judgments in assessing the validity of the evaluation. To do that, the author relied on the principles of the general logic of the assessment (Stufflebeam & Coryn, 2014). The data analysis evidenced the presence of the first principle of the evaluation logic, i.e., the determination of criteria, although the criteria

selection approach has had serious limitations, as shown earlier. The author has identified some problematic issues regarding the second principle, application of standards of quality. Evaluative conclusions have been limited to stating the fact of the achievement or non-achievement of program goals; or labeling programs as effective/non-effective or efficient/non-efficient, mainly based on the assessment of targets. Evidence suggests that no attempt has been made to set gradation or ranking to judge the performance of programs.

Utilizing a single cut-off level of performance (for instance, effective/noneffective), can hardly be described as good practice (Scriven, 1995; Davidson, 2005). The approach taken in the program evaluations does not provide a complete picture of the performance of programs and does not allow for explicitly evaluative conclusions.

Another problematic point is the difficulty of determining and justifying the cut-off level. For instance, as shown in Table 4, some district programs in Zhambyl region have been evaluated as efficient mainly owing to the highest percentage of achieved indicators, while others fell short of their targets and have been found inefficient (unpublished analytical papers of state bodies). In this regard, a reasonable question may arise as to whether the programs that achieved less than 100% of their indicators performed badly or why programs with over 90% of their targets met cannot be considered efficient. The lack of explicit reasoning and justification of the cut-off score seriously weakens the validity and credibility of conclusions.

As for the fourth element of evaluative reasoning - synthesizing performance results to make an overall judgment - the analysis has shown that the program evaluations have simply reported findings on all evaluative components, including the assessment of goal achievement and implementation of legislative standards. For instance, the evaluation of the regional program of Zharminsk district (unpublished analytical papers of state bodies) concludes that there had been ineffective use of budget funds, ineffective planning, non-achievement of some indicators, non-compliance with standards of developing programs, but no attempt was made to weigh and synthesize evaluation findings. Given that some aspects of performance may be of less significance, it is essential to synthesize findings to draw overall evaluative claims (Davidson, 2005).

Table 4 - Evaluation of the efficiency of district programs in Zhambyl region

District development programs in Zhambyl region	1.	2.	Calculation of the indicator of efficiency $\frac{\text{result from column 1}}{\text{result from column 2}} * 100$	Conclusion
	$\frac{\text{overall number of indicators}}{\text{number of achieved indicators}} * 100$	$\frac{\text{overall amount of allocated funds}}{\text{amount of utilized funds}} * 100$		
Bayzak	100%	100%	100%	Efficient
Zhambyl	100%	100%	100%	
Sarysu	100%	100.5%	99.5%	
Zhualy	95,8%	100%	95.8%	Not efficient
Korday	91.3%	101.2%	90.2%	
Moyinkum	95.4%	100%	95.6%	
Talass	95.8%	99.9%	95.8%	

Note - Compiled by the author based on data (unpublished analytical papers of state bodies)

Thus, the overall reasoning pattern observed across program evaluations has entailed only one element of the logic of evaluation (Fournier, 1995), i.e., the determination of criteria, while the remaining three principles (developing standards, measuring performance, synthesizing data to make evaluative judgments) are not reflected in the evaluations. This is quite a disturbing message since basing decisions on dichotomy of effective/non-effective or efficient/non-efficient, and failing to frame conclusions in ‘the vocabulary of grading, ranking or scoring’ (Martens, 2018, p. 27) and adequately synthesizing them, provides only a crude understanding of a program’s value (Scriven, 1995).

It should also be noted that evaluation methodologies and standards employed in Kazakhstan do not address the principles of the general evaluation logic either - apart from identifying criteria. They are more concerned with describing technical and administrative aspects of performing evaluation (for instance, procedures for communicating between evaluators and other state bodies, requirements for documenting reports, and calculation of indicators) or outlining general principles of conducting evaluation (principles of confidentiality, independence, and others), rather than offering specific guidance or strategies for analyzing programs and making evaluative judgments (Adilet, 2020).

Reliability

Ensuring reliability is “a cornerstone for validity” of an evaluation (JCSEE, 2011, p.179). Reliability addresses the consistency and stability of findings and can be achieved through triangulation of data sources and research methods (Golafshani, 2003).

It has already been noted that program evaluations have relied predominantly on statistical records to assess programs and that, in the absence of relevant statistical data, certain aspects of programs were left uninspected. This contrasts sharply with established good practice of evaluation, which calls for the use of various sources, including direct observation and theoretical logical, analogical, or judgmental sources (Scriven, 1991). The fact that information is dependent upon only one type of data does not allow for a comprehensive understanding of the performance of programs and lessens the validity of interpretations (JCSEE, 2011).

In all fairness, it is worth mentioning evaluations that have relied on surveys when assessing the quality of programs. Three of them used surveys among the youth population to determine the level of satisfaction in respect of state measures to support youth (unpublished analytical papers of state bodies). However, the lack of information about the design, procedures, participants, and funding of the surveys does not contribute to judgments about the reliability of the evaluation conclusions. The third evaluation was aimed at determining the impact of health promotion activities in Nur-Sultan city (unpublished analytical papers of state bodies). It was evaluated

on the increase in the number of people practicing a healthy style. It provided a good description of the survey's processes, design, and tools, conducted within the evaluation. Although the survey simply demonstrated percentages of different age groups who practice some form of healthy living without building any causal links, it formed the basis of the report conclusions. Given the occurrence of the effects of such a campaign may take time, other methods, such as interrupted times series, could be used to collect data and make more defensible evaluation findings.

Thus, the analysis has found that there are serious issues which reduce the reliability of evaluation results since these have predominantly relied upon a single source of data; furthermore, apart from in a few cases, they have not utilized any research methods.

Causation

One of the significant components of validity in evaluation is building causal inferences (Davidson, 2005). The causation issue essentially implies determining whether a program has been at least a significant cause of effects or outcomes. It has been found that, although the evaluation standards used in Kazakhstan highlight the need to identify factors and reasons that have affected the realization of a program (Adilet, 2020), establishing causal links has not been a common practice. Numerous evaluations evidence this. The evaluation of the Healthcare program of Karaganda region (unpublished analytical papers of state bodies) concluded that, as a result of the program, mortality rate from tuberculosis was reduced by 34.8% and malignant neoplasm by 7.8%, while infectious disease incidence rate was maintained. However, the lack of explanation on how the program, specifically, caused those changes significantly diminishes the validity of such claims.

There have been only a few programs that have gone further and attempted to draw links between program and observed changes. For instance, the evaluation of the Healthy lifestyle promotion program of Nur-Sultan city (unpublished analytical papers of state bodies) has calculated a correlation between the volume of program activities and the incidence of circulatory system disease. The evaluators have acknowledged the limitations of this technique, as those diseases might be caused by multifaceted factors, and justified this decision by the absence of official statistical data on factors contributing to those incidences. However, the evaluation could have produced more defensible conclusions by applying alternative strategies, such

as asking impacts and observers about the impact of a program, examining if the timing of program effects makes sense, and several others (Davidson, 2005).

The study also identified some evaluations which appear to have shown a clear impact of a program. However, a careful examination shows that inferring causation is still required. It is best illustrated in the evaluation of the Entrepreneurship program of Karaganda region (unpublished analytical papers of state bodies), which assessed a project for organizing a six-month paid internship for new graduates so they could gain initial professional experience and obtain full-time jobs. It has been shown that 867 participants out of 1,047 (83% out of 100%) got hired after completing an internship. However, without developing causal links, it is difficult to credibly argue that it was the program that helped the interns to succeed in getting a job because, potentially, the participants might have done due to their personal skills, background education, and other factors outside the program. To determine this, evaluators could, for example, have interviewed the program participants and asked them explicitly whether the program was the leading cause of their successful employment and in what way it helped them to achieve that if the answer was affirmative.

When addressing the causation issue, it is also essential to look at 'rival explanations' (Davidson, 2005, p. 70), i.e., alternative causes. Sources of such explanations may be found in the context of a program. Contextual factors or other parallel programs may either diminish or enhance the program's effects (JCSEE, 2011). The author has found that this principle has not been practiced in regional program evaluations. Overall, the fact that investigating causal relationships is largely unpracticed when evaluating regional programs is probably one of the most significant limitations of the evaluation practice.

Efficiency and cost-effectiveness

While costs in a broader sense are not limited to monetary costs and, for instance, including human resources, time, and training (JCSEE, 2011), the Kazakhstani regional evaluations have only considered financial costs. Therefore, the author could examine this dimension only.

It has been observed that the determination of efficiency has been based on the examination of program targets and public funds allocated to it. Staying within the approved budget and meeting the targets have been adjudged indicative of an efficient program, whereas failing to stay within the approved budget and/or meet

targets with the same resources demonstrates an inefficient program.

Another consideration when evaluating the efficiency of programs was if program costs were used for the intended purposes and were consistent with principles of the Budget Code and the law, in general. This is another indication of a blurring of lines between evaluation and audit since assessing the compliance of execution of programs with the legislature has traditionally been the prerogative of the audit function.

It is argued that this approach does not wholly correspond with established good practice. Firstly, the judgments about the efficiency of programs are limited in scope and clearly have not been made in a truly evaluative way, as it remains unclear whether the costs were inexpensive, reasonable, or high. Simply noting that costs were efficient/inefficient, based on achievement/non-achievement of goals and proportion of utilized funds is not enough because such claims are insufficiently robust and can be easily subjected to criticism (Davidson, 2005).

To illustrate this, one may pay attention to the evaluation of the regional development program of Saran city in Karaganda region (unpublished analytical papers of state bodies), which was evaluated as 'efficient' based on the fact that 99.8% of indicators were achieved, while the costs stayed within the budget and amounted for 99.8% of the total budget. However, as discussed in the previous section, the evaluations have not dealt realistically with the causation issue, which does not give grounds for asserting with certainty that the program was the primary cause of changes. It might be the case that some criteria of the program (unpublished analytical papers of state bodies), such as a decline in infant mortality, were achieved mainly by the influence of factors outside its scope. Therefore, the claims about the efficiency of program costs, inferred from the fact that indicators were met, are not sufficiently strong and robust. The validity of claims could be significantly substantiated if evaluations included cost-effectiveness or cost-benefit analysis.

Utility

The analysis of evaluations, particularly the Conclusions and Recommendations sections, has revealed several significant findings. First, it was found that recommendations are primarily concerned with the technical and legal aspects of programs. For instance, some evaluations have documented violations of financial planning and budgeting procedures, indicated the need to hold accountable those responsible for mistakes made and provided legal training for servants

to strengthen financial discipline (unpublished analytical papers of state bodies). Other recommendations emphasize that programs must be (i) brought into line with program development standards by developing measurable indicators and clarifying objectives; and (ii) better aligned with strategic and state-level programs. These are meaningful suggestions since they may help to prevent violations that would disrupt the implementation of programs. However, such recommendations have little to do with the substantive content of programs. For example, evaluations have not adequately addressed how the functioning of programs could be improved or what aspects could be modified to achieve better outcomes.

Another set of recommendations refers to the issue of enhancing the overall effectiveness of programs. Nevertheless, most of them are confined to very general statements. For example, program implementers have been encouraged to strengthen coordination between public organizations, enhance monitoring and control over realization of program activities (unpublished analytical papers of state bodies), take measures to improve the efficiency and productivity of programs, etc. It is important to note that these suggestions are already found in the legislature (Adilet, 2020); therefore, program stakeholders would benefit much more from specific, tailored, and actionable recommendations.

Clearly, it is not feasible to assess the implications of the evaluations due to the unavailability of data related to the ultimate use of evaluation findings. However, based on the analysis, the author assumes that the functionality and usability of evaluations in terms of the facilitation of decisions-making is limited, as what has been recommended dealt with technical dimensions of programs or has been too vaguely stated to make adequate use of. Despite this, it is worth mentioning some good practices which may contribute to the application of evaluations. Audit commissions have adopted continuous monitoring of the implementation of recommendations. This procedure is described in detail and enshrined by the rules for conducting audits (Adilet, 2020). Although it is difficult to judge how the cooperation between evaluators and other stakeholders takes place within this practice since the issue has not been reflected in evaluations, the very existence of such practice shows that evaluation does not end with handing over a report and that evaluators are, in fact, willing to provide post-report help (Scriven, 1995). This is vital, as evaluations may require additional explanation and program stakeholders

may have questions or encounter difficulties in the utilization of evaluation findings.

Based on the analysis of the utility component, the author assumes that the usability of evaluation findings in terms of helping in decision-making would be very limited, since there have not been clear and specific conclusions regarding the modification, termination, or enlargement of a program. It is argued that evaluation users could benefit from evaluation results mainly for improving the technical aspects of their programs and bringing them into compliance with legislative norms.

Conclusion and discussion

How well does regional program evaluation conform with established evaluation standards?

The evidence showed that regional program evaluation practice in Kazakhstan has failed to meet all standards applied in this study. Serious discrepancies have been observed both at conceptual and methodological level.

It is understood that Kazakhstan, unlike more developed countries such as the United Kingdom or the United States, does not have a long tradition of program evaluation and that major developments in this sphere took place only in recent years. The author also realizes that evaluators generally act within certain legal constraints and in their work rely on methodologies, which might not be perfect. However, the study found that the scale of the problem is so massive, that it raises questions about a fundamental overhaul of the evaluation practice. To suggest otherwise would be to run the risk of doing a disservice to the Kazakhstani public and contributing to poor decision-making, which may involve considerable sums of taxpayers' money.

To arrive at the conclusions, the author scrutinized regional evaluation practice through the prism of three pillars of good evaluation: values, validity, and utility. The following sections discuss the answers to the research questions.

How justified and appropriate are values used in program evaluation?

The results of the study show that the evaluation reports addressed the values standard very weakly. Firstly, the study found no evidence of attempts to identify and consider all relevant values needed to assess a program. The programs have been assessed from the point of view of their (i) correspondence to indicators set out in the programs; and (ii) compliance with legislative norms (legality of decisions, fulfillment of technical specifications of programs, and legality

of financial costs of programs). The evaluations have not practically considered the values of program recipients and impacts. Another important point left unaddressed was the identification of underlying causes of the performance of programs.

Secondly, the research has shown that program targets have been treated equally without being subjected to scrutiny to determine their relevance and significance; despite the fact that program goals might carry different weights.

Finally, and most importantly, the evaluations have tended to see program targets and legislative norms as intrinsically correct and the sole method of judging outcomes of a program. Furthermore, the evaluations rest largely on the assumption that if targets are achieved and legislative norms are met, it will, inevitably, lead to attainment of program aims and expected results.

How valid are program evaluation design and conclusions?

The research has found several serious issues in this respect which permit the conclusion that the evaluation reports perform very poorly on the validity standard. To demonstrate this, it is worth emphasizing the main findings.

Evaluation logic

The analysis has illustrated that only one of the key principles of evaluative logic has been addressed by the evaluation reports, specifically the identification of criteria. Evidence shows that evaluators have not attempted to set up standards of performance on those criteria in order to state what is weak, good, or excellent performance. Furthermore, the reports do not make clear the evaluative reasoning employed when making claims about a program's effectiveness or ineffectiveness. Finally, the study found that evaluation findings were reported without being weighted and synthesized. The lack of key elements of the evaluation logic gives the grounds to claim that the evaluation reports are not capable of producing explicitly evaluative conclusions.

Reliability

The reliability of the evaluation reports is questionable since they mainly use a limited set of data (official statistical data) to assess programs. This can be explained by the fact that the evaluations were primarily oriented at assessing the achievement of program targets; and the information needed to check that is obtained, as a rule, from official statistics. For fairness, it is worth noting that there has been some use of surveys, but this is the exception rather than the rule.

Causation

Evidence suggests that the practice of establishing causal links in evaluations has been virtually non-existent. This is definitely a serious limitation and evaluation conclusions can hardly be considered valid without addressing the causation issue. This can be illustrated by numerous examples of evaluations attributing changes to the performance of a program without showing logical links between them.

Cost-effectiveness

The concepts of efficiency and cost-effectiveness in the evaluations have been confined to checking if goals were met within a defined budget. This clearly cannot be considered good practice. Firstly, it has already been shown that the program goals might not be valid or justified. Secondly, the assessment of cost-effectiveness cannot be complete without considering alternative ways of spending funds that could produce similar outcomes. The study found no evidence of any tools being employed to achieve this end, such as cost-benefit analysis. Therefore, it is argued that the evaluation reports have performed very poorly on this checkpoint as well.

How useful are program evaluation conclusions and recommendations?

It has been found that the majority of evaluation conclusions and recommendations have dealt with redressing violations of legislative norms, taking measures to prevent them in the future, or improving certain technical aspects of programs and others. Such recommendations can be useful only for making sure that the implementation of programs adhere to norms. However, it is unlikely that evaluation users would be able to make any use of them for improving or changing the content of programs. Meanwhile, there are recommendations related to strengthening the effectiveness of programs, but they are non-specific and therefore not functional. Nevertheless, the author has indicated some good practices; specifically, the practice of reviewing the implementation of evaluation results, which could be an example of post-report interaction and potentially may help evaluation users to apply them.

Practical implications

Based on the research, some policy recommendations can be put forward. Firstly, the understanding of evaluation needs to be conceptually reviewed. We have seen throughout the research that over-emphasizing program goals and treating them as a priori true has led to 'tunnel vision' (Youker & Ingraham, 2014). Consequently, the evaluators have failed to see

other values involved and other intended or unintended effects of the programs.

Secondly, a clear distinction between evaluation and audit should be made. It was found that the evaluation practice has inherited many features of the audit function. A significant aspect of the evaluations has been concerned with checking the conformity of programs to legislative norms and identifying any violations. These are important but the evaluation is much more than that (Chelimsky, 1985).

Thirdly, in the light of the research findings, it seems vital to develop single comprehensive guidance on program evaluation, which would address all aspects of evaluation. Today, evaluators are guided by a plethora of methodological documents. This clearly does not contribute to performing an evaluation in a systematic and focused way. More importantly, the evaluation legislation lacks specific techniques and strategies for design and implementation. In this regard, it might be particularly useful to refer to certain specific public program evaluation methodologies. For example, the United Kingdom's HM Treasury's Magenta Book (Open Government, 2020) provides a good example of systematic evaluation guidance for public programs.

Finally, to ensure high-quality and sound evaluations, the audit authorities of Kazakhstan should consider developing and adopting evaluation standards. Presently, evaluation practice in Kazakhstan lacks professional and sound evaluation standards. The standards for evaluation and audit used today can hardly be described as such in the classical sense since they either describe (i) how administrative procedures should be performed (for instance, how reports should be drawn up and submitted); or (ii) principles of conduct, such as independence, confidentiality, transparency, credibility, and objectivity, rather than providing criteria of quality and guidance on how to achieve them. In this context, a good starting point would be to review the Program Evaluation Standards and Key Evaluation Checklist and explore the possibility of their adoption.

References

1. Legal information system of Regulatory Legal Acts of the Republic of Kazakhstan (Adilet), (2020). <https://adilet.zan.kz/>
2. Akorda, (2020). The Address of the President of the President of the Republic of Kazakhstan at the government meeting [akorda.kz]. Official website of the President of Kazakhstan; 2020. [Updated January 24, 2020; cited November 2, 2021]. Available: https://www.akorda.kz/ru/speeches/internal_political_affairs/in_speeches_and_addresses/vystuplenie-prezidenta-kasym-zhomarta-tokaevana-rasshirennom-zasedanii-pravitelstva-respubliki-kazakhstan.
3. Alibekova, B., Akimova, B. & Ahmetova A. (2019). The formation and current state of audit commissions in the system of state audit bodies. *Statistics, accounting and audit*, 1(72). 16-21.
4. Alibekova, G., Tayova, G. & Ilmaliyev, J. (2018). Evaluation issues of support programs of commercialization of scientific development in Kazakhstan. *Complex use of mineral resources*, 4(307). 181-191.
5. Aymagambetov, E. & Tyngisheva, A. (2019). Modern state of the healthcare system in Karaganda region. *Karaganda University Bulletin*, 2, 11-17.
6. Baimbetov, M. (2019). Implementation evaluation of state programs in the Republic of Kazakhstan. *Bulletin of the Omsk Regional Institute*, 3, 43-46.
7. Bopiyeva, Z. & Kazakova, T. (2009a). Kazakhstan's experience in impact evaluation of state programs. *Bulletin of Legislature Institute of the Republic of Kazakhstan*, 2(14), 20-27.
8. Bopiyeva, Z. & Kazakova, T. (2009b). Ways of improving effectiveness of state programs. *Bulletin of the Legislation Institute of the Republic of Kazakhstan*, 4(16). 160-164.
9. Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: some notes on their relationship. *Evaluation Review*, 9(4), 483-503. <https://doi.org/10.1177/0193841X8500900406>.
10. Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage.
11. Dosayeva, A. (2019). Principles, approaches and features of state audit. *Vestnik Bulletin of Turan University*, 4(84), 149-154.
12. Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 1995(68), 15-32. <https://doi.org/10.1002/ev.1017>.
13. Gayfutdinova, V. (2017). Why development programs don't work in Kazakhstan, [updated November 8, 2017; cited December 10, 2021]. Available: https://forbes.kz/process/expertise/pochemu_v_rk_ne_rabotayut_programmyi_razvitiya_ot_zarubejnyh_konsultantov/
14. Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report*, 8(4), 597-607. <https://doi.org/10.46743/2160-3715/2003.1870>
15. Ho, J. Y., & Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. *Bmj*, 362. <https://doi.org/10.1136/bmj.k3622>
16. Joint Committee on Standards for Educational Evaluation (2011). The program evaluation standards: a guide for evaluators and evaluation users.
17. Kaldiyarov, D. & Turgambekova, Z. (2019). Effective methods for designing and evaluation state program in the social sphere. *Modern Science*, 5(1), 107-109.
18. Kari, A. (2015). Methodological problems of implementation evaluation of strategic documents and programs in public management of economy in the Republic of Kazakhstan. *Economics: the strategy and practice*, 3(35).
19. Martens, K. S. (2018). Rubrics in program evaluation. *Evaluation Journal of Australasia*, 18(1), 21-44. <https://doi.org/10.1177/1035719X17753961>
20. Murzaliyeva, A. & Karshalova, A. (2018). State effectiveness audit of implementation of healthcare budget programs. *Economic systems management: electronic journal*, 4(110), 32.
21. Nygmetov, K. (2014). Strategy and program evaluation in the system of state planning in the Republic of Kazakhstan. *Economics and management*, 34(547), 95-102.
22. The Organization for Economic Cooperation and Development (OECD). OECD Ilibrary. (2021) [updated February 13, 2021; cited March 30, 2022]. Available: <https://www.oecd-ilibrary.org/>
23. Open Government, Magenta Book. Central Government guidance on evaluation, (2020). [updated 2020; cited March 30, 2022]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/HMT_Magenta_Book.pdf
24. Pokidayev, D. (2020). Who will be conducting evaluation of effectiveness of state programs [Kto zaymetsya otsenkoy effektivnosti kazhstanskikh gosprogramm], [updated February 7, 2020; cited December 2021]. Available: <https://kursiv.kz/news/obschestvo/2020-02/kto-zaymetsya-ocenкой-effektivnosti-kazhstanskikh-gosprogramm>
25. Pritvorova, T. & Bektleyeva, D. (2017). Social effect evaluation methods in active labour market projects in Kazakhstan. *Herald of Omsk University. Economics Series*, 3(59). 201-210.
26. Rakhmatullayeva, D., Bobkov, V. & Jatkanbayev, E. (2015). Modelling of social effects of direct foreign investments in regions of Kazakhstan. *Economy of Region*, 2(42), 285-300.
27. Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4), 387-419. <https://doi.org/10.1177/109821400202300403>.

28. Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). California: Sage Publications.
29. Scriven, M. (1995). The logic of evaluation and evaluation practice. *New Directions for Evaluation*, 1995(68), 49-70. <https://doi.org/10.1002/ev.1019>.
30. Shakirova, G., Sitenko, D. and Vasa, L. (2019). Effectiveness audit of public budget expenditures in the process of state audit. *Karaganda University Bulletin*, 7562, 263-274.
31. Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications* (Vol. 50). John Wiley & Sons.
32. Vechkinzova, Y. (2008). Evaluation of implementation of state development programs in the Republic of Kazakhstan. *Journal of New Economy*, 1(20), 105-111.
33. Yessimova, S. (2009). Impact evaluation of state programs in the Republic of Kazakhstan. *University Scientific papers*, 3, 257-260.
34. Youker, B. & Ingraham, A. (2014). Goal-free evaluation: an orientation for foundations' evaluations. *The Foundation Review*, 5(4), 51-61. <https://doi.org/10.9707/1944-5660.1182>
7. Bopiyeva, Z. & Kazakova, T. (2009). Kazakhstan's experience in impact evaluation of state programs. *Vestnik Instituta zakonodatelstva Respubliki Kazahstan [Bulletin of Legislature Institute of the Republic of Kazakhstan]*, 2(14), 20-27. (in Russ)
8. Bopiyeva, Z. & Kazakova, T. (2009). Ways of improving effectiveness of state programs]. *Bulletin of the Legislation Institute of the Republic of Kazakhstan [Vestnik Instituta zakonodatelstva Respubliki Kazahstan]*, 4(16), 160-164. (in Russ)
9. Chelimsky, E. (1985). Comparing and contrasting auditing and evaluation: some notes on their relationship. *Evaluation Review*, 9(4), 483-503. <https://doi.org/10.1177/0193841X8500900406>.
10. Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage.
11. Dosayeva, A. (2019). Principles, approaches and features of state audit. *Bulletin of Turan University [Vestnik Universiteta Turan]*, 4(84), 149-154. (in Russ).
12. Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation*, 1995(68), 15-32. <https://doi.org/10.1002/ev.1017>.
13. Gayfutdinova, V. (2017). Why development programs don't work in Kazakhstan [*Pochemu v Kazahstane ne rabotayut programmyi razvitiya*], [updated November 8, 2017; cited December 10, 2021]. Available: https://forbes.kz/process/expertise/pochemu_v_rk_ne_rabotayut_programmyi_razvitiya_ot_zarubejnyih_konsultantov/ (in Russ)
14. Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report*, 8(4), 597-607. <https://doi.org/10.46743/2160-3715/2003.1870>
15. Ho, J. Y., & Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: retrospective observational study. *Bmj*, 362. <https://doi.org/10.1136/bmj.k3622>
16. Joint Committee on Standards for Educational Evaluation. (2011). *The program evaluation standards: a guide for evaluators and evaluation users*.
17. Kaldiyarov, D. & Turgambekova, Z. (2019). Effective methods for designing and evaluation state program in the social sphere. *Modern Science [Sovremennaya nauka]*, 5(1), 107-109. (in Russ)
18. Kari, A. (2015). Methodological problems of implementation evaluation of strategic documents and programs in public management of economy in the Republic of Kazakhstan. *Economics: the strategy and practice [Ekonomika: strategiya i praktika]*, 3(35). (in Russ).
19. Martens, K. S. (2018). Rubrics in program evaluation. *Evaluation Journal of Australasia*, 18(1), 21-44. <https://doi.org/10.1177/1035719X17753961>
20. Murzaliyeva, A. & Karshalova, A. (2018). State effectiveness audit of implementation of healthcare budget programs. *Economic systems management: electronic journal [Upravleniye*

Список литературы (транслитерация)

1. Legal information system of Regulatory Legal Acts of the Republic of Kazakhstan (Adilet), (2020). <https://adilet.zan.kz/>
2. Akorda, (2020). The Address of the President of the President of the Republic of Kazakhstan at the government meeting [akorda.kz]. Official website of the President of Kazakhstan; 2020. [updated January 24, 2020; cited November 2, 2021]. Available: https://www.akorda.kz/ru/speeches/internal_political_affairs/in_speeches_and_addresses/vystuplenie-prezidenta-kasym-zhomarta-tokaeva-na-rasshirenno-zasedanii-pravitelstva-respubliki-kazahstan.
3. Alibekova, B., Akimova, B. & Ahmetova A. (2019). The formation and current state of audit commissions in the system of state audit bodies]. *Statistics, accounting and audit [Statistika, uchet i audit]*, 1(72), 16-21. (in Russ)
4. Alibekova, G., Tayova, G. & Ilmaliyev, J. (2018). Evaluation issues of support programs of commercialization of scientific development in Kazakhstan. *Complex use of mineral resources [Kompleksnoe ispolzovanie mineralnogo syr'ya]*, 4(307), 181-191. (in Russ)
5. Aymagambetov, E. & Tyngisheva, A. (2019). Modern state of the healthcare system in Karaganda region. *Karaganda University Bulletin [Vestnik KarGU]*, 2, 11-17. (in Russ)
6. Baimbetov, M. (2019). Implementation evaluation of state programs in the Republic of Kazakhstan. *Bulletin of the Omsk Regional Institute [Vestnik Omskogo regionalnogo instituta]*, 3, 43-46. (in Russ)

ekonomicheskimi sistemami: elektronnyi jurnal], 4(110), 32. (in Russ)

21. Nygmetov, K. (2014). Strategy and program evaluation in the system of state planning in the Republic of Kazakhstan. *Economics and management [Ekonomika i Upravleniye]*, 34(547), 95-102. (in Russ)

22. The Organization for Economic Cooperation and Development (OECD). OECD ILibrary. (2021) [updated February 13, 2021; cited March 30, 2022]. Available: <https://www.oecd-ilibrary.org/>

23. Open Government, Magenta Book. Central Government guidance on evaluation, (2020). [updated 2020; cited March 30, 2022]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879438/HMT_Magenta_Book.pdf

24. Pokidayev, D. (2020). Who will be conducting evaluation of effectiveness of state programs [*Kto zaymetsya otsenkoy effektivnosti kazhstanskikh gosprogramm*], [updated February 7, 2020; cited December 2021]. Available: <https://kursiv.kz/news/obschestvo/2020-02/kto-zaymetsya-ocenкой-effektivnosti-kazhstanskikh-gosprogramm>. (in Russ)

25. Pritvorova, T. & Bektleyeva, D. (2017). Social effect evaluation methods in active labour market projects in Kazakhstan. *Herald of Omsk University. Economics Series [Vestnik Omskogo universiteta. Seriya Ekonomika]*, 3(59), 201-210. (in Russ)

26. Rakhmatullayeva, D., Bobkov, V. & Jatkanbayev, E. (2015). Modelling of social effects of direct foreign investments in regions of Kazakhstan. *Economy of Region [Ekonomika regionala]*, 2(42), 285-300. (in Russ)

27. Scott-Little, C., Hamann, M. S., & Jurs, S. G. (2002). Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation*, 23(4), 387-419. <https://doi.org/10.1177/109821400202300403>.

28. Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). California: Sage Publications.

29. Scriven, M. (1995). The logic of evaluation and evaluation practice. *New Directions for Evaluation*, 1995(68), 49-70. <https://doi.org/10.1002/ev.1019>.

30. Shakirova, G., Sitenko, D. & Vasa, L. (2019). Effectiveness audit of public budget expenditures in the process of state audit. *Karaganda University Bulletin [Vestnik KarGU]*, 7562, 263-274. (in Russ).

31. Stufflebeam, D. L., & Coryn, C. L. (2014). *Evaluation theory, models, and applications* (Vol. 50). John Wiley & Sons.

32. Vechkinzova, Y. (2008). Evaluation of implementation of state development programs in the Republic of Kazakhstan. *Journal of New Economy [Izvestiya UrGEU]*, 1(20), 105-111. (in Russ)

33. Yessimova, S. (2009). Impact evaluation of state programs in the Republic of Kazakhstan. *University Scientific papers [Universiteski naukovii zapiski]*, 3, 257-260. (in Russ)

34. Youker, B. & Ingraham, A. (2014). Goal-free evaluation: an orientation for foundations' evaluations. *The Foundation Review*, 5(4), 51-61. <https://doi.org/10.9707/1944-5660.1182>

Information about the author

*Nazym N. Battalov – PhD candidate, Al-Farabi Kazakh National University, Kazakhstan, e-mail: nnbattalov@gmail.com, ORCID ID: <https://orcid.org/0000-0002-9012-7966>

Автор туралы мәліметтер

*Батталов Н.Н. – PhD докторанты, Әл-Фараби қазақ ұлттық университеті, Қазақстан, e-mail: nnbattalov@gmail.com, ORCID ID: <https://orcid.org/0000-0002-9012-7966>

Сведения об авторе

*Батталов Н.Н. – докторант PhD, Казахский национальный университет имени аль-Фараби, Казахстан, e-mail: nnbattalov@gmail.com, ORCID ID: <https://orcid.org/0000-0002-9012-7966>