

МРНТИ 14.35.07  
УДК 371.14

<https://doi.org/10.51889/8354.2022.75.92.035>

*М.Ж. Ильясова*

*«Назарбаев Зияткерлік мектептері» дербес білім беру ұйымының  
«Педагогикалық өлшеулер орталығы» филиалы, Нұр-Сұлтан қ., Қазақстан  
e-mail: ilyassova\_m@cpi.nis.edu.kz*

## **ПЕДАГОГ КАДРЛАРДЫҢ БІЛІКТІЛІГІН АРТТЫРУ ШЕҢБЕРІНДЕ БАҒАЛАУ ҚҰРАЛЫНЫҢ ВАЛИДТІЛІГІ МЕН СЕНІМДІЛІГІ**

*Аңдатпа*

Бұл зерттеу педагог кадрлардың біліктілігін арттыру шеңберінде бағалау құралының валидтілігі мен сенімділігін анықтауға бағытталған. Зерттеу кезінде сапалық және сандық зерттеу әдістері қолданылды. Статистикалық талдау үшін «Математика» және «Информатика» пәндері бойынша әзірленген 118 портфолионы бағалау нәтижелері қарастырылды. Лайкерта тәсілімен алынған сауалнамаға 15 эксперт қатысты. Фокус топқа бағалауды жүзеге асыратын эксперттер және оқытуды жүзеге асыратын тренерлер тандалды. Бағалау валидтілік, сенімділік, жеткіліктілік және түпнұсқалық бағалау қағидаларына негізделеді. Педагогтердің портфолиосын бағалау құралының валидтілігі мазмұндық, құрылымдық, критериалдық дәлелдемелер тұрғысында талданды. Бағалау құралының сенімділігі статистикалық талдау негізінде анықталды. Валидті және сенімді бағалау құралдарын әзірлеу кезінде басшылыққа алынатын ұсыныстар жасалды. Зерттеу нәтижелері бағалаушы эксперттер тәжірибесі үшін құнды болып табылады.

**Түйін сөздер:** біліктілік арттыру, бағалау, валидтілік, сенімділік.

*Аннотация*

*М.Ж.Ильясова*

*Филиал автономной организации образования «Назарбаев Интеллектуальные школы»  
«Центр педагогических измерений», г. Нур-Султан, Казахстан*

## **ВАЛИДНОСТЬ И НАДЕЖНОСТЬ ИНСТРУМЕНТА ОЦЕНИВАНИЯ В РАМКАХ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ ПЕДАГОГИЧЕСКИХ КАДРОВ**

Данное исследование направлено на определение валидности и надежности инструмента оценивания в рамках повышения квалификации педагогических кадров. При исследовании использовались качественные и количественные методы исследования. Для статистического анализа были использованы результаты оценивания 118 портфолио, разработанных по предметам «Математика» и «Информатика». В опросе, проведенном с помощью метода Лайкерта, приняли участие 15 экспертов. В фокус группу были выбраны эксперты, осуществляющие оценивание, и тренеры, осуществляющие обучение. Оценивание основывается на принципах валидности, надежности, достаточности и подлинности оценивания. Валидность инструмента оценивания портфолио педагогов анализируется в контексте содержательных, конструктивных, критериальных доказательств. Надежность инструмента определяется на основе статистического анализа. Представлены рекомендации, который основывается при разработке валидных и надежных инструментов оценивания. Результаты исследования является ценным для практики экспертов по оцениванию.

**Ключевые слова:** повышения квалификации, оценивание, валидность, надежность.

*Abstract*

## **VALIDITY AND RELIABILITY OF THE ASSESSMENT TOOL IN THE FRAMEWORK OF PROFESSIONAL DEVELOPMENT OF TEACHING STAFF**

*Ilyassova M.*

*Center for Pedagogical Measurements of AEO "Nazarbayev Intellectual schools", Nur-Sultan, Kazakhstan*

This study is aimed at determining the validity and reliability of the assessment tool within the professional development of teaching staff. Qualitative and quantitative research methods were used in the study. For statistical analysis, the results of the 118 portfolio' assessment developed in mathematics and computer science were considered. 15 experts took part in the survey obtained by the Likert method. The Focus Group consisted of experts, who carry out assessments and trainers who carry out training. The assessment is based on the principles of validity, reliability, sufficiency and authenticity of the assessment. The validity of the teachers' portfolio assessment tool is analyzed in the context of meaningful, constructive, criteria-based evidence. The reliability of the tool is determined based on statistical

analysis. Recommendations are presented, which is based on the development of valid and reliable assessment tools. The results of the study are valuable in the practice of experts in assessment.

**Keywords:** professional development, assessment, validity, reliability.

### **Кіріспе**

Елімізде педагог кадрлардың біліктілігін арттыру шеңберінде курс соңындағы бағалау педагогтердің оқу кезінде қалыптастырған білімін, түсінігін және тәжірибелік дағдыларын қолдана алуын бағалауға бағытталады. Бағалау валидтілік, сенімділік, жеткіліктілік және түпнұсқалық бағалау қағидаларына негізделген. Валидтілік бағалау үдерісінің бағаланатын нәрсеге сәйкестігі, сенімділік бағалау тұрақтылығы, жеткіліктілік дәлелдемелердің толық зерделенуі, түпнұсқалық ұсынылған дәлелдеменің шынайылығы (аутентикалығы) арқылы қамтамасыз етіледі. Бағалау үдерісінде осы қағидаларды негізге алу бағалаудың әділдігін қамтамасыз етеді [1].

Біліктілік арттыру курстары шеңберіндегі жиынтық бағалау қорытындысында курс тыңдаушыларының табысты оқуының дәлелі ретінде сертификатқа ұсынылады. Көпшілік жағдайда бағалаудың бағаланушылар үшін салдары болады [2], сондықтан да бағалаудың валидті және сенімді болуына баса назар аударылады. Бағалау кезінде валидтілік пен сенімділікке әсер ететін қандайда факторлар кездесіп отырады. Курс кезінде педагогтердің әзірлеген портфолиосын бағалау үшін белгіленген бағалау критерийлерін қамтитын проформа қолданылады. Бір жыл ішінде «Есептелмеді» шешімі қабылданған 249 портфолио модерацияға түскен, оларды модерациялау нәтижесінде оның 5%-на «Есептелді» шешімі қабылданған. Осы шешімдер бойынша бағалау проформаларын талдау кейбір компоненттер бойынша қойылған бағалар арасында алшақтықты байқатты. Демек барлық бағалар дәл, нақты деп айта алмаймыз. Әртүрлі бағалаушылардың бағалары неғұрлым дәйекті болса, бағалау соғұрлым сенімдірек болады деп есептеледі [2]. Бағалаудың сенімділігі мен бағалау компонентері бойынша бағалардың нақты болмауында қарама-қайшылық орын алады. Бағалаудың сенімділігіне әсер ететін факторлардың бірі құралдың валидтілігі [3]. Бағалау құралының сенімділігі оның валидтілігінің қажетті шарты болып табылады. Шығарылған шешімнің сенімділігін қамтамасыз ету үшін құралдың валидтілігін, сенімділігін нысанаға алуымыз керек. Педагог кадрлардың біліктілігін арттыру шеңберінде жиынтық бағалауға арналған валидті және сенімді бағалау құралдарын қалай әзірленуі керек деген сұрақ туындайды. Зерттеу мақсаты валидті және сенімді құрал әзірлеуге ықпал ететін шараларды анықтау болып табылды. Зерттеу нәтижесінде бағалау құралдарын әзірлеу әдістемесі жетілдіріледі деп күтіледі. Егер бағалау құралын енгізуден бұрын алдын-ала оның сапасын бағалайтын болсақ, бағалаудың сапалы және бағалау қағидаларына сәйкес жүзеге асуына қол жеткізіледі. Бұл эксперттер командасының бағалау құралын әзірлеу және олардың сапасын бағалау тәжірибелерін жақсартуға үлес қосатыны анық. Бағалау құралдарын әзірлеу және оның валидтілігі мен сенімділігіне арналған бірқатар ғылыми зерттеулер бар. Moskal (2000) бағалау рубрикасының құрылымын ұсынып валидтілік пен сенімділік, Anders Jonsson (2007) эмпирикалық зерттеулерге шолу жасап баллдық рубрикаларды қолдану, Andrade H.G. (2005) рубрикалардың валидтілігі мен сенімділігі, техникалық оқу орындарында оқыту мен бағалау құралы ретінде қолдану мәселелерін қарастырған [2,9,10]. Студенттердің инновациялық күзиреттілігін бағалау рубрикасының валидтілігін, білім беру ресурстарын бағалауға арналған рубриканың валидациясын зерттеуге арналған еңбектер бар [4,10]. Зерттеулер оқушылардың, студенттердің оқуын бағалау құралдарын әзірлеуге, валидтілігі мен сенімділігін анықтауға, мұғалімдер портфолиосының мүмкіндіктеріне арналған, ал біліктілік арттыру курстары шеңберіндегі бағалау құралдарының валидтілігі мен сенімділік мәселесінің қарастырылмауы зерттеу өзектілігін растайды.

### **Зерттеу әдіснамасы**

Зерттеу педагогтердің біліктілігін арттыру курстары шеңберіндегі бағалау үдерісінде жүзеге асырылды. Зерттеу кезінде математика және информатика пәні мұғалімдері әзірлеген портфолионы бағалау деректері жиналды. Зерттеу кезінде бағалау құралының валидтілігі және сенімділігі аспектісіндегі теориялар негізге алынды. «Валидтілік» ұғымы екі аспектіні қамтиды: не өлшенуі керек және ол қаншалықты дәйекті өлшенеді [6]. Валидтілік бағалау нәтижелері негізінде қабылданатын шешім мен интерпретацияның сапасын бағалау, демек бағалау нәтижесінде біз қаншалықты дұрыс қорытынды жасайтынымызды немесе әрекет ететінімізді түсіну ретінде қарастырылады. Бұл тұста бағалау негізінде қабылданатын шешім мен қорытындының негізді болуы аса маңызды болып табылады. Демек, критерийлерге сәйкес бағаның нақты қойылуына, қабылданатын шешімдердің

негізді болуына қол жеткізу, бұл бағалау валидтілігін қамтамасыз етілуін меңзейді. Демек тек ғана нені бағалайтынызға ғана қатысты емес, портфолионы бағалау кезіндегі интерпретация қаншалықты сапалы жасалғанына, қаншалықты баға және шешім негізді шығарылуына мән беріледі.

Тестілеу теориясында валидтілік мазмұндық, конструктивтік, критериалдық болып бөлінеді [7]. Роберт Линн Дэвид Миллар (2005) валидтілікке қатысты мазмұн, конструктивтілік бойынша пікірлерді, критерийлік қатынастарды қарастыруды ұсынады. Бұл үшеуі дәстүлі, ал төртіншісі салдарлы ойлар [8]. Сондай-ақ В. Moskal (2000) бағалау құралының валидтілігін растау үшін дәлелдемелердің үш түрін, атап айтқанда мазмұн, құрылым және критерийі қарастырылатындығын айтады. Біліктілік арттыру курстарындағы жиынтық бағалау құралын әзірлеуде бағдарлама бойынша қандай мазмұн бағаланатыны қарастырылады. Мазмұнға қатысты дәлелдерді қарастыру кезінде бағалау құралы мазмұнының сәйкестігіне ғана емес, тыңдаушыдан курста игерген білімін, түсінігін және тәжірибелік дағдыларын құрал қаншалықты көрсете алатындығына және қаншалықты дәл өлшейтіндігіне айрықша мән берілді.

Конструкцияларды валидтілік дәлелінің бөлігі ретінде қарастырғанда, бағалау нәтижесін белгілі бір құрылымның немесе дизайнның бар екендігі туралы қорытынды жасау үшін қаншалықты қолдануға болатындығы зерттеледі [5]. В Moskal (2000) бағалаудың мақсаты пайымдауды бағалау болса, өнімнің екеуі де (яғни, жауап) және процесс (яғни, түсініктеме) сұралуы және зерттелуі керектігін айтады. Бұл идеялар бағалау құралының валидтілігін мазмұн жағынан ғана қарамай, құрылымдық жағынан қарастырудың маңыздылығын көрсетіп отыр. Критерийлерді қарастыру валидтілік аргументін әзірлеу біздің бағалау нәтижелерімізден бөлек бағалау немесе критерий бойынша жазылған басқа өлшемдермен қаншалықты жақсы салыстырылатынын немесе болжайтынын қарауды қамтиды [5]. Осыған қатысты айтатын болсақ, құралды критерийлерге қатысты қарастырғанда анықталған критерийлер қаншалықты өлшейді, оның түсініктілігі мен тұжырымдалуына мән беріледі. Демек дәлелдерді интерпретациялау мүмкіндігі бар екеніне, жұмыс туралы пайымдау жасай алатындығымызға көз жеткізу маңызды.

Сенімділік бағалау кезіндегі алынған нәтижелердің тұрақтылығын білдіреді. Уақытқа қатысты тұрақтылық, тапсырмалар арасындағы тұрақтылық, баға қоюшылар арасындағы келісім [3]. Бағалаушылардың қойған бағалары арасындағы тұрақтылық бағалау шешімінің сенімділігін көрсетеді. Әртүрлі бағалаушылар бағалау құралын пайдалана отырып, бір жұмысты бірдей критерийлермен бағалағанда, олар бірдей шешім шығару керек. Демек, бағалау қашан, қайда және кіммен жүргізілсе де тәуелсіз нәтижелер бірдей болуы керек. Сенімділікті тест нәтижелерінің дәйектілігі ретінде қарастырады. Тестіленушіде бағаланатын дағдының қандай да бір деңгейі бар болса, ол қайта тестілеу кезінде дағдының сол деңгейін көрсетеді. Бұл тұста сенімділік бар деп айта аламыз. Сонымен қатар, сенімділікке әсер ететін факторлар да кездеседі. Мысалы, педагогтің мазасындануы, оның портфолио әзірлеуге ынтасының болмауы, бағалаушының субъективтілігі, бағалау құралын пайдалану бойынша бірыңғай түсініктің болмауы. Олай болса, бағалау құралы педагогтер мен бағалаушылар үшін түсінікті болуы маңызды. Бағалау құралының түсініктілігі педагогтердің өз портфолиосын критерийлерге сәйкес дайындауына көмектеседі, ал бағалаушыларға әділ шешім шығаруға ықпал етеді. Зерттеулерде бағалаушылар арасындағы сенімділікті, ішкі сенімділікті анықтау кездеседі. Сенімділікті анықтауда келісім пайызы көрсеткіш ретінде пайдаланылды. Сәйкестікті бағалау әр түрлі корреляция коэффициенттерімен өлшенді. Бағалау құралының валидтілігі мен сенімділігін анықтауда Спирмен статистикалық талдауы үшін ® SPSS ® Statistics 23.0. бағдарламасы қолданылды. Статистикалық талдауға математика және информатика пәні бойынша әзірленген 118 портфолионы бағалау нәтижелері алынды. Лайкерта шкаласымен алынған сауалнамаға бағалауды жүзеге асыратын 15 эксперт қатысты. Сауалнама арқылы респонденттердің проформадағы бағалау компоненттерінің түсініктілігі және оның портфолионы бағалау үшін жарамдылығы 1-5 (өте төменнен өте жоғарыға дейін) аралығында бағаланды. Сауалнама нәтижелерін сенімді деп есептеуге болады, өйткені сауалнамаға қатысқан эксперттердің осы бағалау проформасын 3 жылдық қолдану тәжірибесіне негізделеді. Контент талдау үшін модерацияға түскен 34 портфолионы бағалау проформалары, шешім негіздемелері тандалды. Деректердің сенімділігін қамтамасыз ету үшін фокус топта бейресми сұқбат жүргізілді. Фокус топтың негізгі құрамына 6 эксперт шақырылды, оның екеуі бағалау құралдарын әзірлеумен айналысатын және эксперттер тәлімгері. Зерттеу кезінде пікірдің бір жақты болмауын қамтамасыз ету үшін фокус топ құрамына тренерлер қамтылды.

### Зерттеу нәтижелері

Бағалау құралының мазмұндық, құрылымдық валидтілігін талдауға тоқталайық. Педагогтер өз портфолиоларында математика, информатика пәні бойынша жиынтық бағалау тапсырмаларын ұсынады. Педагогтердің курс соңында бөлім бойынша және тоқсан бойынша тапсырмаларды әзірлей алу, тапсырмаларға сараптама жасау дағдылары қалыптасуы тиіс. Сонымен қатар критериалды бағалау жүйесін қолдана алуы күтіледі. Портфолио мазмұны бағдарлама мазмұнына сәйкес. Оқу нәтижелерін бағалау үшін мазмұн жеткілікті, демек педагогтердің оқу нәтижелері ретінде әзірленген жиынтық бағалау тапсырмалары талап етіледі. Пән бойынша тапсырмалар негізінде игерілген дағдыларды бағалау мүмкін, әрі тәжірибелік сипатқа ие. Алайда оқу нәтижелерін негізге алатын болсақ, тапсырмаларға сараптама жасау дағдыларын бағалау қамтылмаған. Портфолиода курс соңында педагогтердің алған білімін қолдана алуын растайтын мазмұн тандалған, бірақ шын мәнісінде педагогтің түсінігін бағалау мүмкіндігі қарастырылмағанын байқаймыз. Демек, портфолио мазмұндық жағынан валидті, бірақ құрылымдық тұрғыда валидтілігі әлсіз. Себебі өнім, яғни тапсырма бар, бірақ тыңдаушының сараптаманы қалай жасайтынын түсінуін не арқылы бағаланатынын қарастырылмағаны анықталды. Портфолионы бағалау проформасының валидтілігін мазмұндық, құрылымдық дәлелдер тұрғысынан қарастырайық. Құрылымға (конструкцияға) қатысты дәлелдер бағалау құралы толығымен және тек ғана осы құрылымды өлшейтінін растайтын дәлелдер болып табылады [2] деген идеяны басшылыққа алатын боламыз. Бағалау проформасындағы бағалау компоненттерінің жарамдылығын бағалау нәтижелері (1-сурет) 18 бағалау компонентінің 15-нің «өте жоғары» жарамдылығы көрсеткіштері 67-92% аралығын көрсетті, түсініктілігі бойынша (2-сурет) 16 компоненттің көрсеткіштері жоғары.

Шәкіл	Максаттар			Бағалау критерийлері					Тапсырмалар				Дескрипторлар			Тапсырмалар				Балл қою схемасы		
	Максаттар оқу бағдарламасының болып өткен сәйкес	Бағалау критерийлері оқу мақсаттарына сәйкес	Бағалау критерийлері оқу дағдылары деңгейіне сәйкес	Тапсырмалар бағалау критерийлеріне сәйкес	Тапсырмалар оқу дағдылары деңгейіне сәйкес	Тапсырмалар жас ерекшеліктеріне сәйкес	Тапсырмалар анық және түсінікті құрастырылған	Жауапты айқындайтын белгілерсіз құрылған	Жұмысты орындау уақыты рационалды белгіленген	Тапсырмаларды орындау жолдарының бақылауға және өлшеуге болатын	Тапсырманың күрделілік деңгейіне балл саны сәйкес келеді	Тапсырмалар спецификациевен сәйкес	Тапсырмалар мақсуды оқу мақсаттарына сәйкес	Тапсырмалар мақсуды оқу дағдылары деңгейіне сәйкес	Тапсырмаларды орындау уақыты рационалды түрде белгіленген	Жылытық бағалау тапсырмаларының сипаттауы ұсынылған	Балл қою схемасының құрылымында барлық маңызды компоненттер сақталған	Тапсырманың күрделілік деңгейіне балл саны сәйкес келеді				
Өте төмен	0%	0%	0%	0%	0%	17%	0%	8%	8%	0%	0%	17%	0%	0%	0%	0%	0%	0%				
Төмен	0%	0%	8%	0%	8%	0%	0%	8%	33%	0%	8%	8%	0%	0%	17%	17%	8%	8%				
Орташа	0%	8%	17%	8%	8%	33%	17%	17%	17%	8%	0%	17%	8%	8%	0%	17%	25%	8%				
Жоғары	8%	0%	17%	0%	8%	8%	17%	8%	25%	0%	17%	33%	8%	0%	17%	8%	17%	8%				
Өте жоғары	92%	92%	58%	92%	75%	42%	67%	58%	8%	92%	75%	25%	83%	92%	67%	58%	50%	75%				

Сурет 1. Бағалау проформасындағы бағалау компоненттерінің жарамдылығын бағалау нәтижелері

Шәкіл	Максаттар			Бағалау критерийлері					Тапсырмалар				Дескрипторлар			Тапсырмалар				Балл қою схемасы		
	Максаттар оқу бағдарламасының болып өткен сәйкес	Бағалау критерийлері оқу мақсаттарына сәйкес	Бағалау критерийлері оқу дағдылары деңгейіне сәйкес	Тапсырмалар бағалау критерийлеріне сәйкес	Тапсырмалар оқу дағдылары деңгейіне сәйкес	Тапсырмалар жас ерекшеліктеріне сәйкес	Тапсырмалар анық және түсінікті құрастырылған	Жауапты айқындайтын белгілерсіз құрылған	Жұмысты орындау уақыты рационалды белгіленген	Тапсырмаларды орындау жолдарының бақылауға және өлшеуге болатын	Тапсырманың күрделілік деңгейіне балл саны сәйкес келеді	Тапсырмалар спецификациевен сәйкес	Тапсырмалар мақсуды оқу мақсаттарына сәйкес	Тапсырмалар мақсуды оқу дағдылары деңгейіне сәйкес	Тапсырмаларды орындау уақыты рационалды түрде белгіленген	Жылытық бағалау тапсырмаларының сипаттауы ұсынылған	Балл қою схемасының құрылымында барлық маңызды компоненттер сақталған	Тапсырманың күрделілік деңгейіне балл саны сәйкес келеді				
Өте төмен	0%	0%	0%	0%	0%	8%	0%	8%	0%	0%	0%	8%	0%	0%	0%	0%	0%	0%				
Төмен	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	8%	8%	0%	0%	8%	8%	8%	8%				
Орташа	0%	0%	0%	0%	0%	17%	8%	17%	25%	8%	0%	25%	8%	8%	17%	25%	17%	8%				
Жоғары	8%	8%	17%	8%	17%	17%	25%	25%	25%	8%	8%	33%	17%	8%	17%	8%	42%	17%				
Өте жоғары	92%	92%	83%	92%	83%	58%	67%	50%	25%	83%	83%	25%	75%	83%	58%	58%	33%	67%				

Сурет 2. Бағалау проформасындағы бағалау компоненттерінің түсініктілігін бағалау нәтижелері

Сонымен қатар 11 компонент бойынша статистикалық талдау нәтижесіндегі корреляция коэффициенттері эксперт және модераторлар бағалары арасында жоғары, орта оң байланыстың бар екенін көрсетті (3-сурет). Демек бұл компоненттер мазмұн жағынан валидті. Ал «Тапсырмалар жас

ерекшелігіне сәйкес» компонентінің «өте жоғары» жарамдылығы - 42%, «орташа» - 33%, түсініктілігі «өте жоғары» - 58%, бағалар арасында байланыс әлсіз,  $r = 0,168$ . Портфолионың А бөлімін бағалайтын «Жұмысты орындау уақыты рационалды белгіленген» компонентінің жарамдылығын 33%, «төмен» деп бағаланған, түсініктілігі орташа, бағалар арасында байланыс әлсіз,  $r = 0,159$ . Бұл деректер проформаның құрылымдық, критерийлік валидтілігінде мәселе бар екенін көрсетті. Ұсынылған құрылымды, демек портфолио тапсырмаларын (өнімін) өлшейтіндігіне күмән туындайды. Шынымен де бұл компонент маңызды еместігін байқатады, себебі біріншіден, жиынтық бағалау тапсырмаларын әзірлеу және сараптау дағдыларын тексермейді, бағалау мақсатына сәйкес емес. Екіншіден БЖҒ министрлігінің инструктивті хатында әрбір сыныпта бөлім бойынша жиынтық бағалау уақыты жалпы белгіленген. Негіздемелерді қарастыру уақыт көрсетілсе - «2», көрсетілмесе - 0 бағаға бағаланғанын анықтатты, мұнда сонымен қатар негіздемелердің 90%-да компонент атауын жазу арқылы негіздеме жасаған. Бұл мәселені фокус топтағы тренерлер де атап отыр. Компоненттің пайдалануына қатысты сұраққа фокус тобының Х мүшесі компонент бойынша «бір мәнді шешім шығару қиын», ал У мүшесі «компонентті түсінуде қиындық» бар екенін айтады. Компоненттің өлшеуге жарамсыздығын және оның тұжырымдалуының нақты еместігін аңғарамыз. Оның салдарынан бағалаушының жеке интерпретациясына ерік беріледі. Moskal (2000) егер бағалау критерийі бағалау мақсатына сәйкес емес дәлелге шоғырланғанса, онда бағалаушыны дұрыс емес интерпретацияға итермелейді деген пікірді айтады. Дұрыс емес интерпретация бағаның сенімді болмауына алып келетіні анық.

Рo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0,343	,566*	,495**	,471**	,391**	,694**	,616**	,590**	,505**	,576**	,571**	-,372*	-,553**	-,532**	-,409**	-,375**	-,437**	-,416**
2	,366*	,691**	,485**	,289**	,360**	,442**	,470**	,312**	,365**	,274**	,164**	-,289**	-,222**	-,238**	-,123**	-,256**	-,253**	-,193**
3	,137**	,420**	,257**	,123**	,186**	,255**	,179**	,183**	,138**	,269**	,132**	-,060**	-,001**	-,056**	,035**	-,075**	,007**	,078**
4	-,029**	,319**	,209**	,803**	,375**	,043**	-,010**	,065**	-,007**	,029**	,092**	-,201**	-,146**	-,253**	-,123**	-,043**	-,184**	-,217**
5	,031**	,301**	,298**	,482**	,488**	,086**	,066**	,117**	,131**	,125**	,210**	-,257**	-,263**	-,366**	-,153**	-,119**	-,232**	-,287**
6	,276**	,180**	,023**	,210**	-,194**	,158**	,126**	-,028**	-,004**	,068**	-,040**	-,092**	-,098**	-,127**	-,017**	,014**	-,445**	-,353**
7	,066**	,218**	,021**	,119**	,092**	-,073**	-,081**	-,065**	-,200**	-,276**	-,315**	,080**	,098**	,069**	,111**	,016**	-,123**	-,112**
8	,214**	,213**	-,010**	,152**	,166**	,060**	,037**	-,115**	,007**	-,059**	-,069**	,074**	,007**	-,019**	,047**	,069**	-,281**	-,220**
9	,289**	,389**	,189**	,285**	,353**	,223**	,189**	,195**	-,159**	,088**	,073**	,005**	-,064**	-,085**	-,015**	,007**	-,270**	-,173**
10	,234**	,386**	,148**	,204**	,240**	,158**	,190**	,129**	,205**	,094**	,066**	-,208**	-,215**	-,211**	-,120**	-,100**	-,162**	-,114**
11	,295**	,477**	,319**	,362**	,337**	,211**	,091**	,194**	,245**	,087**	-,144**	-,148**	-,173**	-,212**	-,065**	-,005**	-,126**	-,096**
12	-,397**	-,155**	-,142**	-,374**	-,275**	-,455**	-,493**	-,282**	-,377**	-,538**	-,552**	,349**	,733**	,761**	,808**	,854**	,482**	,596**
13	-,454**	-,335**	-,332**	-,394**	-,401**	-,445**	-,440**	-,430**	-,517**	-,348**	-,512**	,822**	,348**	,720**	,813**	,578**	,397**	,550**
14	-,385**	-,209**	-,292**	-,424**	-,353**	-,405**	-,442**	-,361**	-,499**	-,439**	-,539**	,694**	,902**	,781**	,888**	,573**	,373**	,541**
15	-,313**	-,094**	,031**	-,384**	-,284**	-,330**	-,444**	-,363**	-,218**	-,455**	-,502**	,687**	,552**	,699**	,824**	,649**	,544**	,708**
16	-,216**	-,134**	-,244**	-,306**	-,296**	-,289**	-,288**	-,204**	-,186**	-,351**	-,418**	,575**	,528**	,545**	,697**	,599**	,358**	,501**
17	-,339**	-,279**	-,312**	-,461**	-,428**	-,299**	-,338**	-,199**	-,240**	-,326**	-,465**	,524**	,500**	,474**	,694**	,612**	,525**	,573**
18	-,315**	-,185**	-,225**	-,505**	-,388**	-,277**	-,307**	-,171**	-,207**	-,380**	-,517**	,494**	,465**	,443**	,599**	,583**	,531**	,579**

\*. Корреляция значима на уровне 0,05 (двухсторонняя).  
 \*\*. Корреляция значима на уровне 0,01 (двухсторонняя).

Сурет 3. Эксперт және модераторлар бағаларының корреляция көрсеткіштері

В бөлімі бойынша «Тапсырмалар спецификациямен сәйкес» компонентінің жарамдылығы «жоғары» - 33%, алайда «өте төмен» -17%, бағалар арасында байланыс жоғары оң,  $r = 0,849$ . Сол сияқты «Жиынтық бағалау тапсырмаларының сипаттамасы ұсынылған» компоненті жарамдылығы 50% «өте жоғары», алайда 25% «орташа», түсінікті, бағалар арасында байланыс орташа оң,  $r = 0,599$ . Бағалау проформасының бұл компоненттері бойынша мазмұндық, конструктивтілік валидті деуге болады. Дегенмен жарамдылығы «өте төмен» деп бағалаған респонденттер бар. Бағалау бойынша шығарылған бірнеше шешімдер негізделерінде жиынтық бағалау тапсырмаларының сипаттамасы ұсынылмаған

жағдайда, тапсырмалардың спецификацияға сәйкес еместігі туралы жазады немесе керісінше. Демек бұл компоненттердің функциясы бірін-бірін қайталайды. Сондай-ақ бұл компоненттер бойынша бағамдамалық пайымдау жасау мүмкін еместігін негіздемелер көрсетіп отыр. Мұны фокус топтың X мүшесінің «бұл компоненттерге иә немесе жоқ деп жауап беруге болады» деген пікірі де растайды. Сипаттама ұсынылуға қатысты компонент дағдыны бағалаудан гөрі талаптың орындалуын бағалайды. Демек бұл бағалау компоненттері критерийлерге қатысты валидті емес.

Бағалау құралының сенімділігін қарастырайық. Бағалаушылар арасындағы сенімділік екі немесе одан да көп эксперттер арасындағы келісім дәрежесін білдіреді. Ал ішкі сенімділік бір эксперттің бағаларының бірізділігі. Ішкі сенімділік бағалау кезінде сыртқы факторлардың осы эксперттің қалай бағалайтынына әсер етуі деп түсінеміз. Спирмен статистикалық талдау нәтижелеріне зер салайық (3-кесте). Бағалалар арасындағы байланыс жалпы алғанда сенімділікті көрсетеді. Жоғары оң байланыс  $0,691 \leq r \leq 0,843$  аралығында, орташа оң байланыс  $0,488 \leq r \leq 0,599$  аралығында. Дегенмен жоғарыда аталған компоненттер бойынша байланыс әлсіз болып отыр. Проформаларда бұл компоненттер бойынша эксперт жоғары баға қойса, модератор төмен немесе керісінше қойылған. Бұл құралдағы осы бағалау компоненттерінің сенімділігінің әлсіздігін көрсетеді. Құралдың валидтілігін талдау кезінде осы компоненттердің нақты тұжырымдалмауы және кейбір компоненттердің жарамсыздығы анықталды. Құралдың сенімділігіне әсер ететін фактордың бірі – рубриканың тұжырымдалуы деп көрсетілген [3]. Олай бұл тұста осы компоненттердің дұрыс тұжырымдалмау және мақсатқа сәйкес еместігі сенімділіктің әлсіздігіне алып келіп отыр деуге болады. В. Moskal (2000) жақсы құрастырылған рубрикалар бағалаушылар арасындағы айырмашылықтарды толығымен жоймаса да, осы сәйкессіздіктердің пайда болуын азайта алатындығын айтады. Олай болса бағалау құралдарын сапалы әзірлеу бағаның сенімділігіне де ықпал ете алатындығын байқаймыз. Осы бағалар айырмашылығы болған портфолиоардың негіздемелерін талдауда компоненттерге сәйкес бағамдамалық пайымдаулардың берілмеуі, атап айқанда «ішінара» сөзінің жиі кедесуі, бағалау компоненттерінің қайталануы байқалды. Бұл тыңдаушыға өз жұмысын жетілдіруге көмектеспейді және бағалау компоненттерін бағамдамалық пайымдау түрінде жазудың сәтсіз екенін көрсетті. Мысалы, «сәйкес» орнына «сәйкестігі» деп берілген дұрыс еді. Сәйкестігі десе, сәйкестігі туралы пікірлер жазылған болар еді. Жиынтық пікір болғандықтан «сәйкес», «сәкес емес» өте көп орын алады. Олай болса деңгейлер, яғни бағалар бойынша сипаттамалардың қажеттігін білдіреді. Ал сипатталарды беру бағалаушылар бағаларының арасындағы сәйкессіздіктердің пайда болуын азайтады. Соған қарағанда бұл да бағалау құралын әзірлеудегі мәселе болып отыр. Фокус топтың A, D, Z мүшелері баға сипаттамаларының болуы интерпретацияға көмектесетінін, шешім шығарудың жеңілдейтінін айтады. Дұрыс интерпретация болса, субъективтілікке жол берілмейді. Субъективтілік те бағалар арасындағы байланыстың әлсіздігінің себебі болуы мүмкін. Өйткені бұған осы бағалау компоненттерінің жарамдылығы, түсініктілігін бағалауда орта көрсеткіштер болған еді (1,2 кестелер). Демек бағалау құралдарының дұрыс әзірленуі бағалаушылар арасындағы және ішкі сенімділіктің жоғарылауына ықпал етеді деген тұжырым жасалады.

Бағалаушылар аналитикалық рубриканы қолданғанда мейлінше төмен балл беретіндігін анықтаған [9]. Сондай-ақ зерттеу нәтижелерінде холистикалық рубриканың сенімді, ал аналитикалық рубриканы пайдалану одан да сенімді екенін дәлелденген [10]. Біздің жағдайымыздағы проформа аналитикалық рубрикаға жатады, мұнда бағалау компоненттері әрбір бөлім үшін тағайындалған, әрі көп. Сондықтан да бағаның қатаң болуы, яғни төмен болуы бағаның сенімді болмауына алып келеді. Проформаларды талдауда модератор бағасына қарағанда эксперттердің бастапқы бағалауындағы бағалар мейлінше жоғары немесе төмен болғандығы анықталды. Фокус топтағы барлық мүшелер осы бағдарлама бойынша бағалау топтары көп болған кезде, бағалау уақытының аздығын айтты. Бұл факторлар ішкі сенімдікке әсер еткен. Бағалау құралдарын әзірлеуде құралдың жеңіл пайдалануын ескеру үшін маңызды.

### Талқылау

Бағалау құралын әзірлеуде оның тек ғана жиынтық бағалауда қолдануы ғана емес қалыптастырушы бағалау үшін де валидті және сенімді болуын қарастыру маңызды, бағалау құралының дамытушылық сипатына ие болуына мән беру керек. Олай болса, жиынтық бағалау құралдарының қалыптастырушы бағалау үшін пайдалану мәселесі зерттеудің келесі бағыты болып табылады. Бағалау құралының валидтілігін және сенімділігін анықтау оның әлсіз тұстарын анықтатып, оның себептерін жоюға ықпал ететін ұсыныстарды тұжырымдауға алып келді.

Біліктілік артыру курстары шеңберіндегі бағалау құралдарын әзірлеуде келесі ұсыныстар басшылыққа алу ұсынылады: 1) конструктивті валидті болуы үшін бағаланатын құрылымды нақтылап алу; 2) критерийлік тұрғыда валидті болуы үшін бағалау компоненттерін өлшемді тұжырымдау және бағамдамалық пайымдауды қолданбау; 3) бағалау құралының сенімді болуы үшін деңгейлер, шәкілдер, бағалар бойынша сипаттамалар беру, олардың жиынтық негіздеу пікір беруден гөрі қалыптастырушы пікір беруге лайықты етіп сипаттау; 4) бағалау құралында компоненттер арасындағы байланысты тексеру; 5) ішкі сенімділікке әсер ететін факторларға жол бермеу жолдарын ескеру. Бұл ұсыныстарды екперттер командасы бағалау құралдарын әзірлеуде пайдалана алады.

### Қорытынды

Біліктілік арттыру курстары шеңберіндегі бағалау құралы жалпы валидті және сенімді. Дегенмен конструктивтілік, критерийлерге қатысты валидтілік әлсіз. Зерттеу нәтижесінде бағалау құралының валидтілігін қарастыру бағалау құралындағы олқылықтарды анықтатты, құралдың валидтілігін анықтап алу бағалау құралына түзетулер енгізуге жол ашады. Демек бағалау құралын әзірлеуде мазмұндық, құрылымдық және критерийлерге қатысты дәлелдер тұрғысынан қарастыруды шара ретінде енгізу керек. Бұл валидті және сенімді құрал әзірлеудің қадамдарының бірі болып табылады.

### Пайдаланылған әдебиеттер тізімі:

- 1 «Назарбаев Зияткерлік мектептері» дербес білім беру ұйымы «Педагогикалық өлшеулер орталығы» филиалының бағалау рәсімдері, ДББҰ Басқармасының 2020 жылдың 4 қыркүйектегі (№42 хаттамасы) енгізілген өзгертулер мен толықтырулар, 2020, 3-4 б.
- 2 Barbara M. Moskal, Jon A. Leydens (2000). *Scoring Rubric Development: Validity and Reliability*. 2000. DOI <https://doi.org/10.7275/q7rm-gg74>
- 3 Charles Darr (2005). *A hitchhiker's guide to reliability*. 2005. DOI: <https://doi.org/10.18296/set.0623>
- 4 Frances Watts1, Juan A. Marin-Garcia2, Amparo García Carbonell1, Lourdes E. Aznar Mas (2012). *Validation of a rubric to assess innovation competence*. 2012. Том 3 №1. <https://doi.org/10.4995/wpom.v3i1.1159>
- 5 Charles Darr (2005). *A hitchhiker's guide to validity*. 2005. DOI: <https://doi.org/10.18296/set.0639>
- 6 R. Ebel and D.A. Frisbie. *Essentials of Educational Measurement*. (5th Ed.). Englewood Cliffs, N.J.: Prentice-Hall. 1991. University of Iowa.
- 7 Майоров А.Н. *Теория и практика создания тестов для системы образования*. М.: «Интеллект-центр», 2001. 296 с.
- 8 Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching (9th ed.)*. New Jersey: Pearson Education.
- 9 Thikra K. Ghalib1 & Abdulghani A. Al-Hattami (2015). *Holistic versus Analytic Evaluation of EFL Writing: A Case Study*. Online Published: June 29, 2015. <http://dx.doi.org/10.5539/elt.v8n7p225>
- 10 Anders Jonsson, Gunilla Svingby (2007). *The use of scoring rubrics: Reliability, validity and educational consequences*. *Educational Research Review*. 2 (2007) p. 130–144.
- 11 Black P. (1997). *Testing Friend or foe?* London: Falmer Press. <https://doi.org/10.4324/9780203137840>

### References:

- 1 *The assessment procedures of the branch «Center for pedagogical measurements» under autonomous educational organization «Nazarbayev Intellectual schools», amendments and additions to the AEO Management Board dated by September 4, 2020 (Protocol No. 42), 2020, 3-4. (in Kazakh)*
- 2 Barbara M. Moskal, Jon A. Leydens (2000). *Scoring Rubric Development: Validity and Reliability*. 2000. DOI <https://doi.org/10.7275/q7rm-gg74>
- 3 Charles Darr (2005). *A hitchhiker's guide to reliability*. 2005. DOI: <https://doi.org/10.18296/set.0623>
- 4 Frances Watts1, Juan A. Marin-Garcia2, Amparo García Carbonell1, Lourdes E. Aznar Mas (2012). *Validation of a rubric to assess innovation competence*. 2012. Том 3 №1. <https://doi.org/10.4995/wpom.v3i1.1159>
- 5 Charles Darr (2005). *A hitchhiker's guide to validity*. 2005. DOI: <https://doi.org/10.18296/set.0639>
- 6 R. Ebel and D.A. Frisbie. *Essentials of Educational Measurement*. (5th Ed.). Englewood Cliffs, N.J.: Prentice-Hall. 1991. University of Iowa.
- 7 *Theory and practice of creating tests for the education system*. – М.: "Intelligence Center", 2001. - 296 p.(in Russian)
- 8 Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching (9th ed.)*. New Jersey: Pearson Education.
- 9 Thikra K. Ghalib1 & Abdulghani A. Al-Hattami (2015). *Holistic versus Analytic Evaluation of EFL Writing: A Case Study*. Online Published: June 29, 2015. <http://dx.doi.org/10.5539/elt.v8n7p225>
- 10 Anders Jonsson, Gunilla Svingby (2007). *The use of scoring rubrics: Reliability, validity and educational consequences*. *Educational Research Review*. 2 (2007) p. 130–144.
- 11 Black P. (1997). *Testing Friend or foe?* London: Falmer Press. <https://doi.org/10.4324/9780203137840>