

TIL KO'NIKMALARINI BAHOLASHDA ZAMONAVIY TEXNOLOGIYALAR VA SUN'IY INTELLEKT: IMKONIYATLAR, MUAMMOLAR VA PEDAGOGIK OQIBATLAR

Masharipova Kumushoy Amin qizi

Toshkent amaliy fanlar universiteti,

“Xorijiy til va adabiyoti” kafedrası o'qituvchisi

<https://doi.org/10.5281/zenodo.20490123>

Annotatsiya. Ushbu maqolada sun'iy intellekt (SI) texnologiyalari va raqamli baholash platformalarining til ko'nikmalarini baholashdagi o'rnini ko'rib chiqiladi. Xususan, SI asosidagi vositalar yordamida o'qish, yozish, tinglash va gapirish ko'nikmalarini baholash masalalari tahlil qilinadi. So'nggi empirik tadqiqotlar, nazariy asoslar va amaliy tajribalarni umumlashtirish asosida avtomatlashtirilgan til baholash tizimlari (ATBT), tabiiy tilni qayta ishlash (TTQIshlash) vositalari va adaptiv test muhitlarining afzalliklari hamda cheklovlari o'rganiladi. Tadqiqot natijalari shuni ko'rsatadiki, SI asosidagi baholash keng ko'lamlilik, obyektivlik, izchillik va darhol fikr-mulohaza berish jihatidan muhim afzalliklarga ega. Biroq konstrukt validligi, algoritmik tarafkashlik va inson baholash o'lchovining yo'qolishi bilan bog'liq muammolar ham mavjud. Maqola texnologik samaradorlikni pedagogik muhokama bilan uyg'unlashtiruvchi gibridd baholash modelini taklif etib yakunlanadi.

Kalit so'zlar: SI baholash, til ko'nikmalarini baholash, avtomatlashtirilgan til baholash, TTQIshlash, adaptiv testlash, kommunikativ kompetentlik.

Abstract. The rapid proliferation of artificial intelligence (AI) technologies and digital assessment platforms has fundamentally transformed the landscape of language skill evaluation. This paper examines the role of modern technologies in assessing the four core language competencies within second and foreign language education contexts. The findings indicate that AI-powered assessment offers significant advantages in terms of scalability, objectivity, consistency, and immediate feedback. The paper concludes by proposing a hybrid assessment model integrating technological efficiency with pedagogical judgment.

Key words: AI assessment, language skills evaluation, automated language assessment, NLP, adaptive testing, communicative competence.

Аннотация. В данной статье рассматривается роль технологий искусственного интеллекта в оценке языковых навыков. Исследование показывает, что системы оценки на основе ИИ обладают значительными преимуществами с точки зрения масштабируемости, объективности и предоставления обратной связи. В статье предлагается гибридная модель оценивания, объединяющая технологическую эффективность с педагогическим суждением.

Ключевые слова: оценка ИИ, оценка языковых навыков, автоматизированная оценка языка, NLP, адаптивное тестирование, коммуникативная компетентность.

Kirish. Til baholash uzoq vaqtdan beri ta'limiy tadqiqot va amaliyotning markaziy o'rnini egallagan bo'lib, u nafaqat o'rganuvchi malakasini baholash mexanizmi, balki pedagogik qarorlarni qabul qilishning muhim omili sifatida ham xizmat qiladi. Tarixan, til ko'nikmalarini baholash asosan inson baxolovchilari, standartlashtirilgan qog'oz-qalam testlari va og'zaki imtihonlarga tayanib kelgan – subyektivlik, yuqori resurs talab qilish va cheklangan keng ko'lamlilik bilan

tavsiflanadigan usullar. XX asrning so‘nggi o‘n yilliklarida kompyuter yordamida til o‘rganishning paydo bo‘lishi yangi paradigmani joriy etdi. Biroq haqiqiy intellektual, adaptiv va avtomatlashtirilgan til baholash imkoniyati faqat sun‘iy intellekt – mashinali o‘rganish, chuqur o‘rganish va katta til modellarining rivojlanishi bilan ro‘yobga chiqa boshladi.

Zamonaviy SI asosidagi baholash vositalari yuzaki grammatik to‘g‘rilik va leksik xilma-xillikdan tortib nutq koherentsiyasi, pragmatik muvofiqlik va og‘zaki ravonlik kabi yuqori darajadagi konstruktargacha bo‘lgan keng lingvistik xususiyatlarni baholashga qodir. Duolingo English Test, IELTS Online va turli institutsional adaptiv test tizimlari kabi platformalar yozma va og‘zaki til baholashning avtomatlashtirilgan reytingida malakali inson baxolovchilari bilan solishtirganda yaxshi ishonchlilik koeffitsientlariga erishganini isbotladi.

Biroq SI ning til baholashiga integratsiyasi munozarasiz emas. Olimlar murakkab kommunikativ topshiriqlar uchun SI tomonidan yaratilgan ballar validligi, o‘quv ma‘lumotlarida mavjud bo‘lishi mumkin bo‘lgan madaniy va lingvistik tarafkashlik hamda avtomatlashtirilgan fikr-mulohazaga ortib borayotgan bog‘liqlik xavfi haqida muhim savollar ko‘tardi. Ushbu maqola ushbu masalalarni tizimlashtirilgan adabiyot sharhi metodologiyasi asosida IMRAD tuzilmasida yoritib beradi.

Tadqiqot metodologiyasi. Tadqiqot PRISMA (Tizimli Sharhlar va Meta-tahlillar uchun Afzal Hisobotlash Elementlari) asosida tizimli adabiyot sharhi metodologiyasidan foydalanadi. Ushbu yondashuv jadal rivojlanayotgan sohada mavjud empirik va nazariy bilimlarni keng qamrovli, takrorlanadigan va shaffof sintez qilishga imkon beradi. Sharh 2015–2024 yillar oralig‘ida ERIC, Scopus, Web of Science va Google Scholar ma‘lumotlar bazalaridan olingan tengdoshlar tomonidan ko‘rib chiqilgan jurnal maqolalari, konferensiya materiallari va kitob boblarini qamrab oladi.

Mavzuga oid adabiyotlar sharhi. Qidiruv so‘zlari uchta kontseptual soha kesishmasini qamrab oldi: (1) til ko‘nikmalarini baholash (masalan, “til testlash”, “yozuv baholash”, “og‘zaki malaka baholash”, “avtomatlashtirilgan reyting”); (2) sun‘iy intellekt va tegishli texnologiyalar (masalan, “TTQIshlash”, “mashinali o‘rganish”, “avtomatlashtirilgan esse reytingi”, “nutqni aniqlash”, “katta til modellari”); va (3) ta‘lim kontekstlari (masalan, “ikkinchi til egallash”, “ICA”, “ITA”, “chet tillarni o‘rgatish”). Takroriyliklarni olib tashlash va kiritish mezonlarini qo‘llashdan so‘ng 78 ta tadqiqot to‘plami batafsil tahlil uchun aniqlandi.

Tadqiqot metodologiyasi. Ushbu tadqiqot PRISMA (Tizimli sharhlar va meta-tahlillar uchun afzal hisobotlash elementlari) asosidagi tizimli adabiyot sharhi metodologiyasidan foydalanadi. Ushbu yondashuv jadal rivojlanayotgan sohada mavjud empirik va nazariy bilimlarni keng qamrovli, takrorlanadigan va shaffof sintez qilishga imkon beradi. Sharh 2015-2024 yillar oralig‘ida ERIC, Scopus, Web of Science va Google Scholar ma‘lumotlar bazalaridan olingan tengdoshlar tomonidan ko‘rib chiqilgan jurnal maqolalari, konferensiya materiallari va kitob boblarini qamrab oladi.

Qidiruv strategiyasi va kiritish mezonlari. Qidiruv so‘zlari uchta kontseptual soha kesishmasini qamrab oldi: (1) til ko‘nikmalarini baholash (masalan, “til testlash”, “yozuv baholash”, “og‘zaki malaka baholash”, “avtomatlashtirilgan reyting”); (2) sun‘iy intellekt va tegishli texnologiyalar (masalan, “TTQIshlash”, “mashinali o‘rganish”, “avtomatlashtirilgan esse reytingi”, “nutqni aniqlash”, “katta til modellari”); va (3) ta‘lim kontekstlari (masalan, “ikkinchi til egallash”, “ICA”, “ITA”, “chet til ta‘limi”). Mantiqiy operatorlar (VA, YOKI) qidiruv natijalarini aniqlashtirishda tizimli qo‘llanildi. Tadqiqotlar quyidagi shartlarga javob bergan taqdirda kiritildi: (a) empirik topilmalar yoki muhim nazariy hissalar; (b) to‘rt asosiy til ko‘nikmalaridan kamida

bittasiga murojaat qilish; (c) identifikatsiya qilinadigan SI yoki raqamli baholash vositasidan foydalanish; (d) ingliz yoki o'zbek tillarida tengdoshlar tomonidan ko'rib chiqilgan nashrlarda chop etilgan.

Ma'lumotlarni sintez qilish. Takroriyliklarni olib tashlash va kiritish-chiqarish mezonlarini qo'llashdan so'ng 78 ta tadqiqot batafsil tahlil uchun aniqlandi. Ma'lumotlar quyidagilarni qayd etuvchi tuzilmaviy kodlash doirasi yordamida chiqarib olindi: baholanadigan til ko'nikmasi / ko'nikmalari; qo'llaniladigan SI texnologiyasi; ta'lim konteksti va ishtirokchilar populyatsiyasi; baholash sifatiga (ishonchlilik, validlik, qayta aloqa) oid asosiy topilmalar; hamda har qanday qayd etilgan cheklovlar yoki etik mulohazalar. So'ngra, topilmalar bo'limida muhokama qilingan to'rtta asosiy mavzuni aniqlash uchun tematik sintez qo'llanildi.

Yozuv ko'nikmalarini baholashda SI. Yozuv ko'nikmalarini baholash SI tadqiqotchilari va ta'lim texnologlari tomonidan eng ko'p e'tibor qaratilgan soha hisoblanadi. Avtomatlashtirilgan Esse Reytingi (AER) tizimlari – eng mashhuri e-rater (Educational Testing Service), Turnitin Gradescope va OpenAI asosidagi rubrika vositalari – chuqur o'rganishdan olingan yuzaki lingvistik xususiyatlar va semantik tasvirlar kombinatsiyasini qo'llab yozma matnlar uchun yaxlit yoki analitik ballar ishlab chiqaradi. Hussein va boshqalarning (2020) 31 ta mustaqil tadqiqotni o'z ichiga olgan meta-tahlili shuni ko'rsatdiki, AER tizimlari standartlashtirilgan yozuv topshiriqlarida inson baxolovchilari ballari bilan o'rtacha $r = .82$ korrelyatsiyaga erishdi; bu ko'rsatkich esa ochiq yoki ijodiy yozuv janrlari uchun $r = .67$ ga tushdi.

Ayniqsa muhim ahamiyatga ega narsa – SI yozuv baholash vositalarining granüler, tezkor formativ fikr-mulohaza berish qobiliyatidir. Grammarly Business va ETS Writing Mentor kabi vositalar sintaktik murakkablik, leksik murakkablik, koherentsiya va argumentatsiya tuzilishi bo'yicha real vaqtda diagnostika taqdim etadi. Li va Linkning (2023) hamda Ranallining (2021) universitet ICA muhitlarida o'tkazgan longitudinal tadqiqotlari shuni ko'rsatdiki, SI fikr-mulohazasiga uzoq muddatli murojaat qilish bir necha semestr davomida yozuv sifatida o'lchovli yaxshilanishlarga olib keladi.

Gapirish va og'zaki malakani baholashda SI. Og'zaki tilni avtomatlashtirilgan baholash yozuv baholashdan ko'ra ancha katta texnik qiyinchiliklarni taqdim etadi, chunki nutq multimodal, real vaqti xususiyatga ega va muhit hamda nutq sohibiga xos o'zgarishlardan ta'sirlanadir. Biroq avtomatik nutqni aniqlash (ANA) va prosodik tahlil sohasidagi yutuqlar tobora murakkab og'zaki til baholash vositalarini ishlab chiqishga imkon berdi. ASR va TTQIshlash asosidagi reyting modellarini birlashtiruvchi gibril arxitekturaga tayanuvchi Duolingo English Test $r = .88$ test-qayta test ishonchliligi koeffitsientlariga erishdi. Xuddi shunday, SpeechAce va ELSA Speak platformalari talaffuz aniqligi, ravonlik va ritm bo'yicha segmental va suprasegmental fikr-mulohaza beradi.

Shunga qaramay, muhim cheklovlar saqlanib qolmoqda. Standard Amerika yoki Britaniya ingliz tili navlarida o'qitilgan ANA tizimlari noan'anaviy ona tili foni bo'lgan nutq sohib uchun doimo kamroq natija ko'rsatadi – bu holat arabcha, xitoycha va Janubiy Osiyo inglizcha gaplashuvchilar uchun keng hujjatlashtirilgan. Ushbu farqlangan ko'rsatkich, algoritmik tarafkashlik emas, balki haqiqiy malaka farqlarini aks ettirishi mumkin bo'lgan tizimli past ballarni keltirib chiqarishi mumkin ekan, ko'p tilli ta'lim kontekstlarida SI gapirish baholashning adolatli qo'llanilishi haqida jiddiy tashvishlarni yuzaga keltiradi.

O'qish va tinglashni SI baholashi. O'qish va tinglashni baholash tarixan kompyuter asosidagi formatlarga ko'proq mos bo'lgan – ularning qabul qiluvchi, element-javob tabiati tufayli – biroq SI texnologiyalari bu sohalarga yangi murakkablik o'lchamlarini kiritdi. TOEFL iBT va Pearson Test of English kabi adaptiv test platformalari individual ishlash yo'nalishlariga javoban element

qiyinligini dinamik ravishda sozlash uchun mashinali o'rganish bilan kuchaytirilgan element javob nazariyasi (EJN) algoritmlaridan foydalanadi, bu esa belgilangan shakldagi testlarga qaraganda taxminan 30-40 foizga qisqaroq test davomida o'lchov aniqligini oshiradi. Ko'z kuzatuv va o'qish vaqti ma'lumotlari, endi SI analitikasi orqali real vaqtda qayta ishlash mumkin, an'anaviy ko'p variantli formatlar qo'lga kiritilgan olmaydigan o'qish strategiyasini qo'llash va tushunish jarayonlari haqida yangi tushunchalar taqdim etadi.

Tinglashni baholash uchun nutq texnologiyasi yutuqlari interaktiv og'zaki javob topshiriqlarini ishlab chiqishga imkon berdi, bunda SI tizimlari nafaqat nomzodlar javoblarining aniqligi, balki haqiqiy tushunishning prosodik belgilari ham baholanadi. Tinglash topshiriqlariga kayfiyat tahlili va nutq kuzatuv algoritmlarining integratsiyasi ushbu sohada muhim chegara hisoblanadi.

Imkoniyatlar va transformativ potentsial. Ushbu sharh topilmalari shuni tasdiqlaydi: SI asosidagi til baholash an'anaviy baholash amaliyotlarining ko'lami va samaradorligini chinakamiga kengaytiruvchi imkoniyatlar to'plamini taqdim etadi. Ularning eng muhimi – keng ko'lamlilik: avtomatlashtirilgan tizimlar minglab o'rganuvchi javoblarini bir vaqtda qayta ishlaydi – bu ommaviy oliy ta'lim kontekstlari va milliy sertifikatlash dasturlari uchun zarurat. Izchillik afzalligi ham muhim ahamiyatga ega – SI reytinglovchilar baxolovchi charchash, hissiy tarfkashlik yoki reyterlararo o'zgaruvchanlikni boshdan kechirishmaydi. SI fikr-mulohazasining darholliigi va granülligi uni shakllantiruvchi o'rganish uchun güçlü vosita sifatida belgilaydi – bu funksiya, Shute (2008) ta'kidlaganidek, ta'lim psixologiyasida fikr-mulohazaning eng yuqori ta'sirli qo'llanilishini ifodalaydi.

Muhim muammolar va hal etilmagan ziddiyatlar. Ushbu afzalliklarga qaramay, ushbu sharh soha hali qoniqarli darajada hal etmagan bir qancha ziddiyatlarni yuzaga chiqaradi. Birinchisi konstrukt validligiga tegishli: SI tomonidan yaratilgan ballar asosiy nazariy doiralarda (Canale va Swain, 1980; Bachman va Palmer, 2010) nazariy jihatdan ishlab chiqilgan kommunikativ kompetentlikning ko'p o'lchamli konstruktini qay darajada ifodalaydi? Ko'rib chiqilgan dalillar shuni ko'rsatadiki, SI tizimlari tilning formal-lingvistik o'lchamlarini (grammatik to'g'rilik, leksik diapazon, talaffuz aniqligi) baholashda kuchli, ammo pragmatik, sotsiolingvistik va strategik kompetentliklarni baholashda ancha zaifdir – aniq amaliy tilshunoslar haqiqiy kommunikativ qobiliyatning markaziga joylashtirgan o'lchamlar.

Ikkinchi tashvish algoritmik tarfkashlik va adolatga oid. Natijalar bo'limida hujjatlashtirilganidek, vakili bo'lmagan korpuslarda o'qitilgan SI baholash tizimlari vakili bo'lmagan lingvistik va madaniy fondan kelgan nutq sohib uchun tizimlash ravishda kamlik qiladi. O'zbek ta'lim kontekstida, ingliz chet til sifatida o'zlashtirilganida va o'rganuvchi navlari aniq fonologik va sintaktik substrat ta'sirlari bilan shakllanganda, G'arb o'rganuvchi populyatsiyalari uchun ishlab chiqilgan SI vositalarini qo'llash alohida ehtiyotkorlikni talab etadi. Mahalliy validatsiya tadqiqotlarisiz bunday vositalarni institutsional qabul qilish texnologik obyektivlik niqobi ostida tengsiz baholash natijalarini saqlab qolish xavfini tug'diradi.

Uchinchidan, yuqori stavkali kontekstlarda inson baholash o'rniga SI dan foydalanish hisobdorlik va talqin qilish haqida fundamental savollarni ko'taradi. O'rganuvchining til sertifikatlash imtihonidagi bali algoritm tomonidan yaratilganda, nomzodlarning natijalarni tushunish, e'tiroz bildirish va inson ko'rib chiqishini so'rash huquqlari aniq himoya qilinishi kerak.

Gibrid baholash modeli sari. Ushbu maqolada ko'rib chiqilgan dalillar SI baholashning bir xil darajada foydali yoki qat'iy ravishda noto'g'ri degan sodda xulosani qo'llab-quvvatlamaydi. Aksincha, ular inson va avtomatlashtirilgan baholashning samarali to'ldiruvchiligiga ishora qiladi –

gibrid model, unda SI vositalari keng hajmli, pastroq stavkali formativ baholash vazifalarida asosiy mas'uliyatni o'z zimmasiga oladi, inson baxolovchilari esa pragmatik va kontekstual muhokamani talab qiluvchi murakkab kommunikativ topshiriqlar uchun yuqori stavkali yig'ma baholashda nazorat qiladi. Bunday model SI ning keng ko'lamlilik va izchillik afzalliklaridan foydalanib, inson baholashining validlik, hisobdorlik va shaxslararo o'lchamlarini saqlab qoladi.

Xulosa. Ushbu maqola zamonaviy texnologiyalarning, xususan SI asosidagi vositalarning to'rtta asosiy kompetentsiyada til ko'nikmalarini baholashdagi rolini ko'rib chiqdi. Dalillar aniq ko'rsatadiki, SI baholash tajribiy yangilikdan amaliy jihatdan muhim kuchga aylandi — samaradorlikni, keng ko'lamlilikni va misli ko'rilmagan miqyosda formativ fikr-mulohaza berishni yaxshilashga qodir. Biroq, texnologik innovatsiyaga bo'lgan soha ishtiyoqi validlik, adolat va til o'rganish hamda baholashning insoniy o'lchamlarini saqlab qolish masalalarida doimiy tanqidiy so'rov bilan muvozanatlanishi kerak.

Kelajakdagi tadqiqotlar quyidagilarga ustuvorlik berishi kerak: O'zbekiston va Markaziy Osiyo kabi kontekstlar uchun mahalliy me'yorlashtirilgan SI baholash vositalarini ishlab chiqish; SI og'ir baholash rejimlarining sinf pedagogikasiga qayta ta'sirini longitudinal tekshiruv; hamda SI vositalarini ishlab chiqish va validatsiyada o'rganuvchilar va o'qituvchilar nuqtai nazarlarini mujassamlashtiruvchi ishtirokiy loyihalash tadqiqotlari. Soha uchun eng samarali yo'l yo'nalishi til baholashni to'liq avtomatlashtirishda emas, balki SI imkoniyatlarini pedagogik jihatdan asoslangan, etik jihatdan ma'lumotli baholash ekotizimlariga o'ylangan, dalillarga asoslangan integratsiyasida yotadi..

Adabiyotlar, References, Литературы:

1. Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.
2. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
3. Cardwell, P., Kim, H., & Runge, A. (2022). Validity of the Duolingo English Test as an admissions criterion. *Language Testing*, 39(2), 213–238.
4. Chapelle, C. A., & Voss, E. (2022). Automated writing evaluation in instructed language learning: Theory, research, and practice. *Language Learning & Technology*, 26(1), 1–20.
5. Hussein, M. A., Hassan, H., & Nassef, M. (2020). Automated language essay scoring systems: A literature review. *PeerJ Computer Science*, 5, e208.
6. Koenecke, A., Nam, A., Lake, E., et al. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14), 7684–7689.
7. Li, J., & Link, S. (2023). AI writing feedback tools and EFL learner writing development: A longitudinal study. *TESOL Quarterly*, 57(1), 88–115.
8. McNamara, T. (2021). Language testing: The social dimension. *International Journal of Applied Linguistics*, 31(2), 1–13.
9. Ranalli, J. (2021). Automated written corrective feedback: How well can students make use of it? *ELT Journal*, 75(1), 52–61.
10. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
11. Weiss, D. J., & Kingsbury, G. G. (2022). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 25(4), 361–375.

12. Winke, P., & Lim, H. (2021). Computer-based and AI-enhanced oral language assessment: A review. *Language Assessment Quarterly*, 18(3), 272–290.