MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2023, 19(9), em2319
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/13471

The geographic inquiry process skills scale: A validation study

Bagila Sadykova [1*], Ulzhalgas Yesnazarova [1], Aigul Tokbergenova [1]

[1] Department of Geography, Land Management, and Cadastre, Al-Farabi Kazakh National University, Almaty, KAZAKHSTAN

Received 30 April 2023 ▪ Accepted 03 July 2023

Abstract

Geographic inquiry has immense potential to spark the interest of school students in science and societal issues, such as climate change or resource scarcity. However, implementing inquiry-based learning in secondary school contexts is frequently seen as a challenge. So far, standardized geography assessments have primarily focused on students' spatial-thinking abilities, and there is a dearth of practicable tools to measure their inquiry skills. This study aimed to translate the self-report geographic inquiry process skills scale into Kazakh and test its reliability and validity among Kazakhstani secondary students. A total of 826 secondary school students aged between 13 and 18 were included in the analysis. Both exploratory and confirmatory factor analyses jointly supported a five-dimensional structure of the questionnaire. The scale exhibited sound measurement properties, including consistency over a two-week test-retest interval. The scores for the adapted instrument were not significantly correlated with participant gender, grade, age, or time spent preparing for the cross-national geography assessment. Proposals for future research are outlined.

Keywords: geography, Kazakhstan, scale adaptation, secondary school, secondary students

INTRODUCTION

Inquiry, Inquiry Skills, and Inquiry Learning

The rapid and incessant advancements achieved by science and technology since the mid-twentieth century demand more and more competent, creative, and independent people who can put new ideas and technological innovation at the service of social progress. Paradoxically, this scientific and technological development has also led to knowledge becoming obsolete at an accelerated pace, hence the need in individuals who can take ownership of knowledge. Therefore, it is the responsibility of educational systems to equip the population with the capacity to autonomously attain updated knowledge necessary to make the appropriate decisions and select the pertinent alternatives that will allow individuals to effectively solve personal, professional and social problems. In fulfilling this commitment, pedagogical approaches ought to be implemented that facilitate cognitive independence to enable self-directed learning as a means of constant training. In particular, the shift from teacher-centered to student-centered educational strategies has been trumpeted as a desideratum across the globe again and again (Eze, 2021; Rafiq et al., 2023). One such strategy is inquiry-based curriculum designed to nurture students' inquiry skills. Inquiry-based learning is a student-centered instructional model grounded in Jean Piaget's constructivist theory of learning, which posits that knowledge is actively constructed by students, is incorporated through instruments of study, and causes the students to become conscious and responsible actors of their learning (Bidell & Fischer, 1992).

There are myriad of interpretations of the inquiry skills construct in literature. Skills related to performing of inquiry mirror those inherent in scientific process skills like posing questions, making inferences, identifying patterns, and so forth. Kuhn and Pease (2008) outline the skills of inquiry as the capability to identify an addressable question (the causal role of a specific feature), seek out informative data through controlled comparisons, and obtain accurate conclusions about causal relationships. They view inquiry not as the accumulation of facts, but rather as an activity that progresses by agglutinating findings with the theoretical landscape. When formulating a definition of inquiry

✉ bagilas@yahoo.com **(*Correspondence)** ✉ ataulymektep@mail.ru ✉ tokbergen@mail.ru

Contribution to the literature

- The adaptation of the geographic inquiry process skills scale in a non-English-speaking setting is beneficial for global geography education.
- Owing to this study, most Central Asian populations can now use a reliable scale to measure self-efficacy in geographic inquiry among secondary students.
- This study provides new evidence for geography education research. This study contributes to geographic inquiry, which has immense potential to spark the interest of school students in science and societal issues, such as climate change or resource scarcity.

skills, Wu and Hsieh (2006) put explanation building at the core, which in turn demands intellectual abilities such as an individual's capacity to coordinate their own exploratory procedures and manipulate variables or features, e.g., causal variables affecting a phenomenon. According to Pedaste et al. (2012), inquiry involves students formulating hypotheses, testing them through trials, and summarizing their espials. This implies the development of transferable abilities compulsory for making new discoveries. Learning to ask probing questions to grasp nature of the world is pivotal during the process of maturation. Early indications of this cognitive function can even be observed in early childhood. As compiled in Lehtinen et al. (2022), the abilities required to investigate the connections between two variables, form hypotheses, and navigate self-learning typically evolve between the ages of five and ten. However, older children and teenagers often face challenges in inquiry, allegedly because it involves orchestrating several cognitive processes like generating hypotheses, selecting the most adequate ones, making observations, and revising viewpoints. This can be illustrated by the so-called active sensemaking loop described, inter alia, in Kachergis et al. (2017), which somewhat mimics the circuit of scientific reasoning aimed to gain scientific knowledge (Kremer et al., 2013). Developmental constraints can be imposed by inefficiencies in that chain. For instance, a person may successfully gather information to test hypotheses, but if they fail to appropriately revise their ideas against the mined evidence, then the inquiry cycle is disrupted. Students often struggle to construct a hypothesis, devise and execute a sound experiment, interpret evidence, and make inferences (Baum, 2013; Kapici et al., 2019; Nehring et al., 2015). If learners have not acquired skills such as formal reasoning, the mode of content delivery makes little difference since they are unlikely to digest scientific ideas (Towne, 2009).

Generally, the goal of conventional thinking training is to get learners used to make informed choices and solve issues, with most courses putting emphasis on looking into the content in order to leap from the problem to decisions, not on actually performing inquiries. Meanwhile, the act of inquiry is primarily concentrated on pinpointing the issue, circumscribing its concept, and weighing the pros and cons of potential solutions (Darmuki et al., 2023; Preskill & Torres, 1999). Even though inquiry-based learning is problem-focused, digging out a clue to the puzzle is not the foremost concern there. Instead, the accent lies on the inquisition phase as well as the learners' initiative to delineate the problem, consider it, and arrive at a solution (Hussien et al., 2021). Furthermore, "taking action," as argued by Gillon and Stotter (2011), is a crucial extra stage, whereby learners are expected to not only acquire new knowledge but also apply it through developing something new or sharing the knowledge. Evaluation is another important step, where learners review their work and draw lessons to be ready to make beneficial adjustments and advancements in their later inquiry. Inquiry is thus a self-directed and iterative process in which students must acquire matter on their own through questioning and problem-solving (Schultz & DeMers, 2020). Besides, it is individual-specific. Within the scientific discovery as dual search model elaborated by Klahr and Dunbar (1988), inquiry-based learning is conceptualized as a quest between two interacting issue compartments: hypothesis and experiment domains. The former comprises all the variables in a research area and conceivable links between them, whereas the latter encompasses all the experimentations that may take place in that domain. The model connotes two key knowledge sources of the hypothesis space are prior subject knowledge and outcomes from investigations; this points up two alternative strategies to assist students in expressing hypotheses (Kuang et al., 2022). Ergo, while some individuals enter into inquiry through the path of hypothesis, others lack the knowledge and abilities needed to contrive a valid hypothesis, so they tend to start from an experiment that could yield a hypothesis (Lehtinen et al., 2022).

Inquiry skills can be seen as lifelong learning skills that enable individuals to learn meaningfully and permanently by researching and questioning. Thanks to inquiry, not only do learners build up knowledge, but they also get a profound comprehension of scientific principles, which can be fruitful for them beyond the classroom (Perugini & Bodzin, 2020). It is believed that individuals who engage in inquiry are likely to develop better abilities to formulate questions, systematize their thoughts, design and carry out experiments, deduce, and communicate their findings (Aldahmash & Omar, 2021).

This bespeaks inquiry is a germane educational topic. Numerous standards worldwide recognize the cultivation of students' competencies related to partaking in and reflecting on inquiry activities as a central mission of education (Petermann & Vorholzer, 2022). For example, in the United States, the science learning performance criteria (outlined in the next generation science standards) are implemented in schools to articulate expectations for learning outcomes and provide educators with the flexibility to set up classroom experiences that spark students' interest towards science and prepare them for academia, employment, and citizenship. Interestingly, nearly all of the practices described in these standards align with the scientific inquiry cycle.

Geographic Inquiry Skills: Problem Statement and Rationale for This Research

The educational research community has long been inclined (though not unanimously) to favor the inquiry-based pedagogical approach over conventional instruction in terms of learner academic gains (Al Mamun et al., 2022; Wilson et al., 2009), including the secondary education level (Jeskova et al., 2022; Petridou et al., 2022; Spires et al., 2022). First-order meta-analyses published over the past decade revealed inquiry-based interventions had had medium to large beneficial effects on student learning (Heindl, 2019; Furtak et al., 2012; Lazonder & Harmsen, 2016). By the same token, a recent second-order meta-analysis states a moderate overall effect size for inquiry learning (Ozturk et al., 2022). Geography itself provides linkage between natural and social sciences, and inquiry as a facet of scientific literacy practice is a critical point in geography education (Jeskova et al., 2022). Geographic inquiry holds tremendous potential to provoke pupils' interest in science and partnerships, along with involvement in world-scale and local sociopolitical problems such as scarcity of resources or climate change (Maddox et al., 2018; Oberle, 2020; Senisum et al., 2022; Xuan et al., 2019). One exemplar is a case study (Schlemper et al., 2018), where American students, building on their own inquiry questions, undertook fieldwork almost throughout their community to gather and synthesize satellite data and observational notes, which resulted in citizen maps highlighting community issues like small and untidy parks, with solutions proposed to the community stakeholders.

Geo-inquiry often entails looking for information on the Internet, and this, alongside the ubiquity of the Internet itself, has spurred the need to foster the Internet-research skills; this has been translated into the piecemeal experimental incorporation of web-based inquiry learning into curricula over the last years (Moreno & Bartolome, 2021), but it has not yet emerged as a full-blown subfield. Nevertheless, the Internet technology has dramatically expanded the range of opportunities for high-quality, scaffolded, and entrancing inquiry experiences in artificial environments like digital atlases, Google expeditions, Esri online applications, and lots more (see, for example, Osborne et al., 2019). This advances data visualization and induces individuals to think more creatively about spatial patterns and interactions between different types of data (Jant et al., 2019). Howbeit, the use of inquiry learning skills by secondary school students is often considered a problematic issue. They were reported to "not consider all relevant variables in a problem and often rely on anthropomorphic explanations in their understandings of the natural world" (Wu & Hsieh, 2006). International testing systems such as the program for international student assessment have recurrently shed light on poor inquiry skills even among European schoolchildren (Ganajova et al., 2021). A look at the issue from a non-student perspective reveals that literature tends to attribute it either to flaws in educational policies, e.g., the lack of administrative support (Curtis, 2019; Oda et al., 2019), or to teachers' imperfections including limited proficiency to use geospatial technologies (Harte & Reitano, 2016; Hohnle et al., 2015; Khalaf & Zin, 2018; Seow et al., 2019).

Policymakers generally rely on findings from standardized assessments to make rational decisions regarding what is best for youth education (Solem et al., 2021). The optimal option here appears to be the objective examination by means of expert-led practical assessments or computer-based tests. Unfortunately, this certainly requires sizable investments in human and economic resources, which is often an elusive goal within the school context, even in countries such as Australia (Timms et al., 2018), not to mention Central Asian countries like Kazakhstan. Therefore, a self-report study is perhaps the most realistic avenue of school student skill evaluation when we talk about the Global South in the foreseeable future. The first trouble that comes to mind with self-rating scales is the risk of the measure skew emanating from the respondent's implicit desire for conforming to norms, so the self-view may represent the person the surveyee would like to be (Brenner & DeLamater, 2016). That said, research evidence offers a rebuttal: Feyzioglu (2019) performed an empirical study underpinned by Bandura's social learning theory (Bandura, 1997) and found that the secondary schooler's inquiry-based self-efficacy level is positively and significantly correlated with his/her level of inquiry skills.

When describing the merits of the self-developed scoring guide for geographic skills, Marcello (2009) remarked that his tool can help children realize how high the standards for geographic learning are. The same can be said for self-administered questionnaires, just as these too, enable teachers to capture the student's background, detect gaps in knowledge and skills, and perhaps even predict his/her potential for geo-inquiry. An overview of experimental research on secondary school geography shows that student assessment revolves mainly around spatial-thinking abilities (González & Torres, 2020; Kim & Bednarz, 2013; Lane & Bourke, 2017; Metoyer & Bednarz, 2016), while there is a paucity of practicable instruments for measuring their inquiry process capacities. Regarding the Kazakhstani sample, a better understanding of the status quo could, inter alia, aid in the development of informed educational strategies aimed at meaningful geography curricula. To date, no well-documented knowledge is available on secondary school students' geographic inquiry skills in this country, be those measured through a self-administered questionnaire or otherwise. With this in view, solid tools are necessary to bridge this gap. Under these circumstances, what we consider a worthy stab at bringing forth is the geographic inquiry process skills scale (GIPSS) recently introduced by Ozudogru and Demiralp (2022) to quantify how students in grades 9 to 12 perceive their level of those skills. The form is unsophisticated and concise, thereby making it possible to keep a lid on the attention span issue at the filling out step. GIPSS has been validated among tenth and eleventh-grade students in Turkey and resultantly demonstrated excellent discriminative power, validity (factor loadings between 0.518 and 0.930), and reliability (Cronbach's alpha between 0.926 and 0.950). Deploying this questionnaire does not entail humungous expenses and is thus adequate for the economic realities in developing countries. Given that such scales are non-existent in the Kazakhstani setting, piloting GIPSS in the local context is definitely of value to Kazakhstan's education system. This could serve as a prospective resource for the Kazakh-language research and a screening self-report instrument for gauging secondary students' geographic inquiry process skills in Kazakhstan. Furthermore, we believe that global academia would also profit from evidence on how replicable this tool is in a non-English and non-Turkish-speaking population. To clarify, the original form is in Turkish, but since the items in the published paper are English-language, we decided to translate them into Kazakh from English rather than Turkish for the sake of transparency. As it is nearly undoable to apply a given measure directly in dissimilar countries due to the language barriers, as well as cultural and social differences, novel scales should be re-examined in various populations to re-examine their practicality (Bhat et al., 2021). In this vein, current validation study could contribute to cross-cultural research on GIPSS.

Research Aim

The purpose of this investigation was to translate, adapt and validate the original GIPSS questionnaire in order to see whether the Kazakh GIPSS can be applied to secondary students in Kazakhstan. Moreover, it sought to explore the adapted GIPSS measurement invariance across grade, age, gender, and a period of preparation for the cross-country geography assessment.

METHOD

Instrument

Upon the validation of GIPSS originally conceived as a 36-item scale comprising six factors, developers cut it down to 22 items measured on a five-point Likert scale ("1=never" to "5=always") distributed between five factors, namely "communicating geographic information," "organizing and analyzing geographic information," "answering geographic questions," "acquiring geographic information," and "posing geographic questions." Respondent perceptions expressed in scores for each item are summed up into a total score. The higher the score, the better geographic inquiry skills an individual is alleged to have.

As the first step, two independently working native Kazakh translators produced two Kazakh-language versions of the scale, which were presented to a four-member expert panel including three geography experts and one pedagogy expert. They examined the clarity of language and integrated the two translations into one form, then back-translated it into English to ensure the correspondence of the items with the original text. To appraise the content validity of the Kazakh draft, i.e., make sure that it gauges what it is supposed to gauge, each expert rated the adequacy of each statement to the intent of the scale and assigned a corresponding value from -1 to 1 to the item, which allowed calculating the item-objective congruence (IOC) index for each item, guided by the concise equation introduced by Turner and Carlson (2003), with 0.75 declared as the minimum cut-off. A pilot experiment was then carried out to see if Kazakhstani school children could accurately understand the prefinal form. The definitive tool was thus acquired and examined as described below.

Participants and Data Collection

School education in Kazakhstan comprises primary (grades 1 to 4), middle (grades 5 to 9), and secondary (grades 10 and 11) levels. In this paper, however, secondary education is mentioned in the general sense, i.e., education that takes place between the ages of 13 and 18. This is a routine practice that as early as at the ninth-grade level Kazakhstani school students start preparing for the national standardized assessment (several-subjects examination held upon the completion of the eleventh grade, mandatory for admission to higher education). This descriptive study was conducted among school students taking the online geography course offered by the app-based learning platform named Joo (joo.kz), which is designed to render preparation assistance with the unified assessment to school students.

In mid-December 2022, the correspondence author used the in-app chat to send invitations to participate in the survey to 1,000 individuals randomly picked from several thousand Joo geography course takers, with clarifications on the survey goal. A total of 904 students agreed and gave informed consent to be involved. In mid-January 2023, the GIPSS final form was made available on the Qualtrics survey system and distributed to the participants by means of the anonymous link option so that the surveyees' identifying information could not be stored. Yet, 50 randomly selected participants were requested to share their responses provided they took the survey, and to fill out the questionnaire again two weeks later to investigate the test-retest reliability of the instrument, that is, whether the responses were subject to substantive fluctuation after a narrow time interval, which would call into question the applicability and reproducibility of the scale. The questionnaire was filled out in a self-reported fashion. Apart from GIPSS items, a few collateral variables were requisite to indicate, namely age, sex, grade, and how long the subject had been preparing for the geography section of the national standardized assessment at the time when she/he filled out the form (less than a year; a year or more; two years or more). In early March 2023, complete replies totaled 849 (response rate of 93.9%) of which 23 were dropped due to the same answer in all items. Thus, 826 valid questionnaires were eventually obtained, and the survey was closed.

The final sample consisted of 826 school students attending the 9th (235, 28.5%), 10th (392, 47.5%), and 11th grades (199, 24.1%), with females slightly prevailing (427, 51.7%). The participants aged between 13 and 18, 15.9 in average (standard deviation [SD]=0.77). Lastly, 568 subjects (68.8%) indicated their period of preparation for the national standardized assessment was <1 year, while 164 (19.9%) reported they spent ≥1 year, and 94 surveyees (11.4%) claimed the period of ≥2 years. It is worth noting that we do not have any information on which schools Joo users attend, so it is impossible to say how many schools this survey embraced.

Data Analysis

All statistical procedures were run in the R software environment. Mean and standard deviation were used to present continuous data. The distribution of each item was checked by means of skewness (a measure of distribution symmetry) and kurtosis (a measure of tail extremity) statistics, with the norm set at the range between -3.0 and 3.0 (Madan & Wang, 2021). To judge whether the entire data acquired from the survey were eligible for factor analysis, the Kaiser-Meyer-Olkin sampling adequacy index along with the Bartlett test of sphericity preceded factor extraction. The raw data was only deemed fit for validation procedures once KMO was higher than 0.60 and the Bartlett test was statistically significant (Gao et al., 2020). Significance was set at levels below $p=0.05$ for all analyses.

Internal consistency reliability

Internal reliability of the translated GIPSS was verified by inspecting floor and ceiling effects and computing corrected item-total correlation, a correlation matrix, Spearman-Brown coefficient, and Cronbach's alpha along with the alpha-if-dropped criterion, which indicates Cronbach's alpha reliability value when a given item is eliminated from the measure. When calculating the percentage of minimum and maximum achievable scores for each item, ceiling effects were inferred to exist if 15.0% of surveyees responded on the item with "always," whereas floor effects were ascertained if 15.0% stated "never." Because the worst and best scorers could not then be differentiated, floor or ceiling effects impair the instrument reliability (Wan Hassan et al., 2017). The corrected item-total correlation was employed to elucidate whether any item possessed insufficient predictive power and therefore was incoherent with the total test score on the other items (Zijlmans et al., 2019), with a coefficient above 0.20 deemed acceptable (Dilbaz et al., 2020). Associations between the subscales were estimated through the Pearson correlation coefficient. Lastly, the internal reliability assessment included the split-half test using an even-odd approach, that is even numbered, and odd-numbered items were split into two subsets and the correlation between them was used to calculate the Spearman-Brown coefficient. The latter is a complementary robust criterion that implies correction for the number of items. Regarding survey scales, Cronbach's alpha from 0.70 to 0.80 is usually regarded as noteworthy and from 0.80 to 0.90 as very good, but when it exceeds 0.90, truncating the form is recommended (Kanbay et al., 2021). The same interpretation scheme was applied to the alpha-if-dropped and split-half method outputs.

Construct validity

The survey responses were divided into two equal parts, one of which ($n=413$) was submitted to exploratory factor analysis (EFA), while the other ($n=413$) to confirmatory factor analysis (CFA), both performed to test the construct validity of the adapted tool. Parallel analysis was selected as a technique for drawing the optimal number of factors from the data matrix (Tak et al., 2022; Zhdanov et al., 2022). Items were retained on the basis of the explained total variance, factor loadings, and a visual inspection of the generated scree plot. Interpretation of conventional scree plots is somewhat intricate and subject to ambiguity as to where exactly the scree lies. With this in mind, the present study resorted to the code written by Sakaluk and Short (2016) to construct a scree plot incorporating parallel analysis. The plot highlights the plausible factor solution by depicting a dashed line across the point following which factors with eigenvalues derived from observed data ceased to outnumber those based on randomly simulated zero-factor matrices. In adherence to common practice, items were seen as admissible for inclusion in a given factor if their loading values were above or equal to 0.40. The quantity of variance in a model that can be explained by the predictor variables rather than measurement error is referred to as the explained total variance (Lam et al., 2005). The extracted factors should account for at least 50% of the total variance (Sarstedt & Mooi, 2014, p. 247). To make the pattern of loadings clearer, the Promax rotation method was applied in EFA and CFA as the simulation evidence shows the robustness of the Promax technique exceeds those of other rotation approaches (Panaretos et al., 2019). Both EFA and CFA were run using the maximum likelihood extraction method, each based on separate data from 413 subjects. A structural equation modeling technique known as CFA (Quinn et al., 2020) was adopted to evaluate the degree of alignment between the measurement model hypothesized within EFA and observed data, with a factor loading of 0.35 as a minimum threshold, in line with Keefer et al. (2019). Model-data fit was examined using the Tucker-Lewis index, comparative fit index, the ratio of chi-squared to the degree of freedom, standardized root mean square residual, and root mean square error of approximation.

Item discrimination statistics

To discover the discriminative capability of each item, the mean scores were compared between individuals with the lowest 27.0% possible scores and those with the highest 27.0% scores using an unpaired two-sample t-test.

Predictive validity

Multivariate regression was utilized to explore whether GIPPS overall and subscale scores were affected by the collateral variables.

External reliability

The test-retest reliability analysis sought to find out the degree of similarity between the outcomes from the same scale administered with a small-time gap. Pursuant to the recommendations outlined in Koo and Li (2016), the intraclass correlation coefficient was estimated based on a two-way single-measurement (because individual sum score rather than average measurement was the variable of interest) mixed-effects model with an absolute agreement to assess concordance between test and retest scores. The coefficient below 0.50 is conventionally interpreted as poor, moderate if it is 0.50-0.75, as good if it is 0.75-0.90, and as excellent if it is over 0.90. Since the evaluation of the test-retest reliability based upon only the intraclass correlation values may lead to misleading conclusions (Avelino et al., 2022), the Bland-Altman method was exploited to determine the test-retest stability of the tool by computing the reproducibility coefficient and plotting the difference between average scores yielded on the first and second survey completions. If a construct is measured twice under identical conditions, the absolute difference between the two measurements will, in 95% of cases, be smaller than the repeatability coefficient, and thus the lower this coefficient is, the greater reproducibility of the survey results (Vaz et al., 2013). A value identifying a tolerable level of mean difference for the Bland-Altman test must be not in excess of one-third of the possible range of scores for the instrument (Ding & Vancleef, 2022). Given that the GIPSS is composed of 22 items with a five-point response option, the potential sum of points lies between 22 and 110. Thus, the acceptable average difference between the two occasions, in this case, should be no more than 29.

RESULTS

Content Validity

As the first step, two independently working native Kazakh translators produced two Kazakh-language versions of the scale, which were presented to a four-member expert panel including one geography expert and three pedagogy experts. They examined the clarity of language and integrated the two translations into one form, then back-translated it into English to ensure the correspondence of the items with the original text. As per the panel and the research team, some minor linguistic adjustments were required to make the questionnaire suitable for the local audience. For instance, item 1 was reduced to "recognize the problem" since "issue" and "problem" are identical semantic constructs in Kazakh. To avoid confusion, "geographic" was added to words like "problem" or "information" in the translated text. In the original study, factor analysis placed the item "being able to select geographic information from data sources" in the subscale accommodating skills related to posing geographic questions. But the item obviously has nothing to do with asking questions, and therefore on a consensus basis was allotted back to the subscale named acquiring geographic information.

Following the adaptation procedures, the experts confirmed the resulting item pool has adequate content validity (IOC of each item ranged from 0.82 to 0.94), so the research team utilized it as a pilot draft in order to assess the scale comprehensibility through cognitive pre-testing. To that end, three high school students outside the research sample volunteered to fill out the preliminary questionnaire. Beyond just selecting numbers, the volunteers were asked to read the items aloud and narrate how they perceive the statements, why they chose given answer, and whether anything

Table 1. Original & translated items in GIPSS

Item	English (original)	Kazakh
CGI1	Determine the fit-for-purpose way	Географиялық ақпарат бойынша қорытындыларды берудің қолайлы әдісін анықтау
CGI2	Use various methods	Географиялық ақпарат бойынша қорытынды берудің әртүрлі әдістерін қолданыңыз
CGI3	Determine target population correctly	Мақсатты аудиторияны дұрыс анықтаңыз
CGI4	Use a clear, effective, & comprehensible way of expression	Ақпаратты жеткізудің тиімді әдістерін қолданыңыз
CGI5	Use geographic terms	Географиялық терминдерді қолданыңыз
OAGI1	Convert geographic information into maps, graphs, & tables	Географиялық ақпаратты карталарға, графиктерге және кестелерге түрлендіру
OAGI2	Read maps & graphs	Карталар мен графиктерді оқыңыз
OAGI3	Interpret maps, graphs, tables, & photos	Карталарды, графиктерді, кестелерді және фотосуреттерді түсіндіріңіз
OAGI4	Make predictions on graphs, tables, & photos	Диаграммалар, кестелер және фотосуреттер бойынша болжамдар жасаңыз
OAGI5	Compare maps, graphs, tables, & photos	Карталарды, графиктерді, кестелерді және фотосуреттерді салыстырыңыз
AGQ1	Reach conclusions	Тұжырымдарды тұжырымдау
AGQ2	Develop thoughts about conclusions	Жасалған тұжырымдарға сүйене отырып ойды дамыту
AGQ3	Make predictions on future based on conclusions	Қорытындылар негізінде болжамдар жасаңыз
AGQ4	Make a decision based on conclusions	Қорытынды негізінде шешім қабылдау
AGI1	Determine path to follow	Географиялық ақпаратты іздеу және іріктеу бағытын анықтау
AGI2	Decide method to be used	Географиялық ақпаратты іздеу және таңдау әдісін анықтаңыз
AGI3	Detect required information	Қажетті географиялық ақпаратты дәл табыңыз
AGI4	Select geographic information from data sources	Дереккөздерден географиялық ақпаратты таңдау
PGQ1	Recognize issue or problem	Бар географиялық мәселені тану
PGQ2	Define issue or problem	Географиялық мәселені тұжырымдау
PGQ3	Distinguish between geographic & non-geographic questions	География бойынша сұрақтарды басқалардан ажыратыңыз
PGQ4	Ask geographic questions	География бойынша сұрақтар қою

Note. For easy reference, items are numbered in a way different from original & abbreviations stand for corresponding factors in scale: CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions

should be amended to improve clarity. Finally, students rated the item pool as understandable, no corrections were considered necessary, and it was approved as the final Kazakh version of GIPSS (hereafter, GIPSS-Kz). **Table 1** gives an overview of the original English wording of the items and the Kazakh translation.

Descriptive Statistics and Internal Consistency Reliability

Internal consistency of the questionnaire was tested on the pooled response rates. Cronbach's coefficient obtained for the 22-item instrument was 0.941 (0.935, 0.947) and it did not surge when any of the items was effaced. The split-half estimate was 0.961 (0.955, 0.966) meaning the data partitioning did not deflate the internal reliability level and it is therefore safe to state that the items measure the same underlying construct. Corrected item-total correlations were moderate. None of the items was scored 1 or 5 by 15.0% of surveyees, implying that neither floor nor ceiling effects were observed in respondents' perceptions, that is no extreme lower or upper part of item pool was lost and thus there was no deterioration of content validity. All skewness values were found close to zero and none of the kurtosis values exceeded 3.0, so the data could be declared normally distributed. Descriptive statistics and internal consistency reliability evidence are given in **Table 2**. As summarized in **Table 3**, there were medium-to-large positive intercorrelations among all GIPSS-Kz subscales, and they all were significant ($p<0.01$).

Construct Validity

EFA

Kaiser-Meyer-Olkin value was 0.95, thus indicative of the adequate sample size, and Bartlett's statistic was found significant (χ^2_{231}=12,271.26; $p<0.01$), altogether implying that numerical data could be subjected to factor analysis procedures.

Table 2. Descriptive statistics & reliability coefficients for the Kazakh version of GIPSS (n=826)

Item	Mean (SD)	Skewness	Kurtosis	Corrected item-total correlation	Alpha if item omitted	Floor (%)	Ceiling (%)
CGI 1	2.99 (1.03)	0.01	2.47	0.433	0.830	7.1	7.1
CGI 2	3.05 (1.01)	-0.12	2.49	0.492	0.827	6.7	6.5
CGI 3	2.98 (0.97)	-0.08	2.61	0.326	0.834	6.7	4.8
CGI 4	3.04 (0.96)	-0.01	2.54	0.360	0.833	4.7	5.8
CGI 5	2.99 (1.05)	-0.06	2.31	0.370	0.832	7.8	6.3
OAGI 1	3.08 (0.94)	-0.01	2.49	0.360	0.833	3.5	5.7
OAGI 2	3.06 (0.94)	-0.17	2.57	0.417	0.831	5.0	4.5
OAGI 3	3.09 (0.87)	0.10	2.49	0.380	0.832	1.5	4.5
OAGI 4	3.12 (0.96)	-0.07	2.55	0.404	0.831	4.0	6.4
OAGI 5	3.10 (0.99)	-0.05	2.42	0.329	0.834	4.5	6.8
AGQ 1	3.14 (0.89)	0.23	2.70	0.441	0.829	1.5	7.3
AGQ 2	3.15 (0.89)	0.13	2.55	0.481	0.828	1.5	6.5
AGQ 3	3.23 (0.91)	-0.02	2.65	0.280	0.837	2.1	7.5
AGQ 4	3.19 (0.90)	0.01	2.82	0.288	0.836	2.5	7.3
AGI 1	3.16 (0.89)	0.18	2.70	0.462	0.828	1.6	7.5
AGI 2	3.16 (0.86)	0.30	2.84	0.442	0.829	1.0	7.5
AGI 3	3.15 (0.91)	0.12	2.67	0.435	0.829	2.2	7.4
AGI 4	3.09 (0.90)	0.23	2.47	0.406	0.831	1.2	6.0
PGQ 1	3.01 (0.95)	0.07	2.53	0.447	0.829	5.6	4.4
PGQ 2	3.03 (0.94)	-0.02	2.58	0.399	0.831	4.4	5.0
PGQ 3	3.05 (0.96)	-0.04	2.59	0.433	0.830	5.0	5.7
PGQ 4	3.02 (1.0)	-0.04	2.43	0.382	0.832	5.8	5.8

Note. Abbreviations stand for corresponding factors in scale: CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; PGQ: Posing geographic questions; & SD: Standard deviation

Table 3. Correlations among factors in the Kazakh version of GIPSS (n=826)

	CGI	OAGI	AGQ	AGI	PGQ
CGI	1.00	0.73	0.35	0.43	0.72
OAGI	0.73	1.00	0.43	0.44	0.75
AGQ	0.35	0.43	1.00	0.64	0.41
AGI	0.43	0.44	0.64	1.00	0.42
PGQ	0.72	0.75	0.41	0.42	1.00

Note All correlations are significant at 0.01 level & abbreviations stand for corresponding factors in scale: CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions

Within EFA (n=413), the maximum likelihood method suggested a five-dimensional structure that accounted for 13.7%, 13.67%, 12.8%, 11.37, and 8.6% of the total variance, with a cumulative variance contribution of 60.2%.

Therewith the parallel analysis demonstrated that a model constituted by five factors could explain more variance in the measure items compared with the synthetic data (**Figure 1**), and thus the five-dimensional solution was pre-accepted, with factor names borrowed from the original tool.

All the items of GIPSS-Kz were entered after the Promax rotation and had factor loading (i.e., the regression of a factor on an item) values ≥0.40 (**Table 4**).

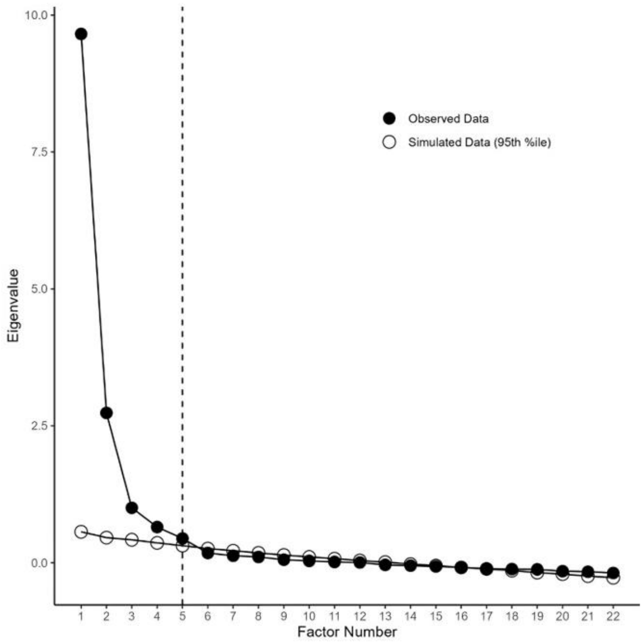

Figure 1. Scree plot of EFA of the Kazakh version of GIPSS (n=413) (Source: Authors' own elaboration)

Table 4. Factor loadings for EFA of the Kazakh version of GIPSS (n=413)

Item	CGI	OAGI	AGQ	AGI	PGQ
CGI 1	0.73				
CGI 2	0.67				
CGI 3	0.78				
CGI 4	0.82				
CGI 5	0.76				

Table 4 (continued). Factor loadings for EFA of the Kazakh version of GIPSS (n=413)

Item	CGI	OAGI	AGQ	AGI	PGQ
OAGI 1		0.59			
OAGI 2		0.81			
OAGI 3		0.92			
OAGI 4		0.72			
OAGI 5		0.73			
AGQ 1			0.87		
AGQ 2			0.77		
AGQ 3			0.93		
AGQ 4			0.72		
AGI 1				0.65	
AGI 2				0.89	
AGI 3				0.92	
AGI 4				0.61	
PGQ 1					0.65
PGQ 2					0.64
PGQ 3					0.69
PGQ 4					0.70

Note. Abbreviations stand for corresponding factors in scale: CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions

CFA

Thereafter, CFA was operated on the other part of the items' responses (n=413) to inspect the measurement model. The analysis proved that the survey data sufficiently fit the a-priori five-factor structure represented by "communicating geographic information," "organizing and analyzing geographic information," "answering geographic questions," "Acquiring geographic Information," and "posing geographic questions" latent variables since model-data fit indices conform the recommended values (**Table 5**).

Table 5. Reference values for goodness-of-fit criteria & observed values of five-factor model of the Kazakh version of GIPSS obtained from CFA (n=413)

GFI	RV	OV (95% CI)	Reference
RMSEA with 90% CI	Below 0.08 (upper bound of CI below 0.10)	0.067 (0.061, 0.074)	Machado et al. (2020)
SRMR	Below 0.10	0.055	De La Torre et al. (2022)
$\chi 2/\text{df}$	Equal or below 5.0	2.866 (570.372/199)	Soares et al. (2022)
CFI	Above 0.90	0.942	Soares et al. (2022)
TLI	Above 0.90	0.932	Soares et al. (2022)

Note. RMSEA: Root mean square error of approximation; CI: Confidence interval; SRMR: Standardized root mean square residual; χ^2/df: Ratio of Chi-squared to degree of freedom; CFI: Comparative fit index; TLI: Tucker-Lewis index; GFI: Goodness-of-fit indicator; RV: Recommended value; OV: Observed value;

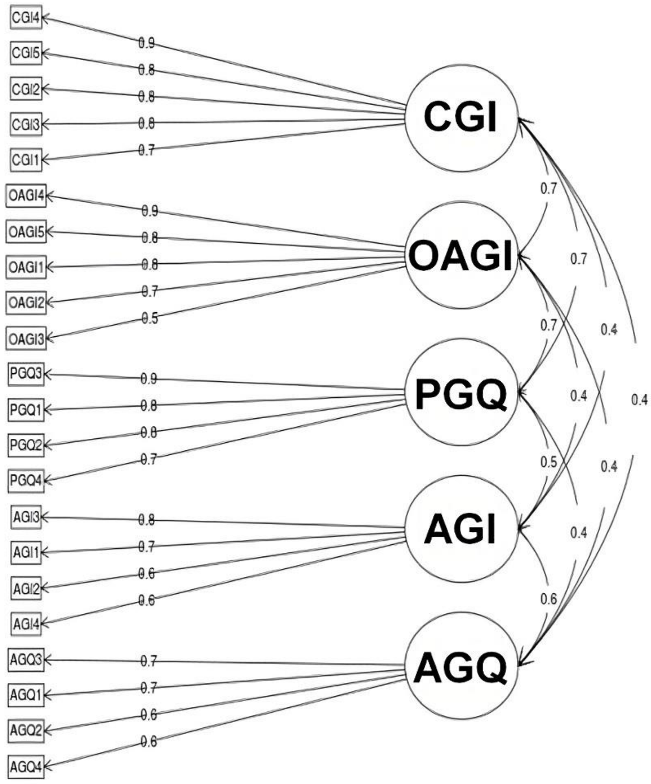

Figure 2. Standardized five-factor structural model of the Kazakh version of GIPSS obtained from CFA (n=413); Rectangles: Measured variables; Circles: Latent (unobserved) factors; One-way arrows: Factor loadings; Two-way arrows: Covariances; CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions (Source: Authors' own elaboration)

Loadings for all 22 items met 0.40 cut-off, which also means the model is supported by the data (**Figure 2**).

Discriminative Validity

Table 6 shows that for each of the items, statistically significant differences were detected between the 27.0% of the total sample who scored the lowest and the 27.0% with the top scores ($p<0.01$). This gives grounds to believe the scale has the capacity to distinguish school students subjectively possessing low and high geographic inquiry process skills.

Table 6. Comparison between 27% of responders with bottom scores (n=223) & those with top 27% scores (n=223) for the Kazakh version of GIPSS (n=826)

Item	High-score group (n=223) mean (SD)	Low-score group (n=223) mean (SD)	t-test (df)	p
CGI1	4.27 (0.44)	1.74 (0.44)	60.408 (444)	<0.01
CGI2	4.24 (0.43)	1.75 (0.43)	61.019 (444)	<0.01
CGI3	4.18 (0.39)	1.75 (0.43)	62.639 (444)	<0.01
CGI4	4.22 (0.41)	1.83 (0.38)	63.632 (444)	<0.01
CGI5	4.23 (0.42)	1.71 (0.45)	60.640 (444)	<0.01

Table 6 (continued). Comparison between 27% of responders with bottom scores (n=223) & those with top 27% scores (n=223) for the Kazakh version of GIPSS (n=826)

Item	High-score group (n=223) mean (SD)	Low-score group (n=223) mean (SD)	t-test (df)	p
OAGI1	4.21 (0.41)	1.87 (0.34)	65.974 (444)	< 0.01
OAGI2	4.17 (0.37)	1.82 (0.39)	65.189 (444)	<0.01
OAGI3	4.17 (0.37)	1.98 (0.29)	69.032 (444)	<0.01
OAGI4	4.24 (0.43)	1.89 (0.42)	58.971 (444)	<0.01
OAGI5	4.25 (0.44)	1.83 (0.37)	63.031 (444)	<0.01
AGQ1	4.27 (0.45)	2.09 (0.43)	52.631 (444)	<0.01
AGQ2	4.24 (0.43)	2.06 (0.41)	54.889 (444)	<0.01
AGQ3	4.28 (0.45)	2.16 (0.54)	45.118 (444)	<0.01
AGQ4	4.27 (0.44)	2.15 (0.57)	43.978 (444)	<0.01
AGI1	4.28 (0.45)	2.12 (0.47)	49.674 (444)	<0.01
AGI2	4.28 (0.45)	2.20 (0.48)	46.982 (444)	<0.01
AGI3	4.27 (0.45)	2.05 (0.46)	51.585 (444)	<0.01
AGI4	4.22 (0.42)	1.96 (0.21)	72.614 (444)	<0.01
PGQ1	4.21 (0.41)	1.84 (0.37)	64.506 (444)	<0.01
PGQ2	4.18 (0.39)	1.84 (0.37)	65.407 (444)	<0.01
PGQ3	4.21 (0.41)	1.82 (0.39)	63.431 (444)	<0.01
PGQ4	4.22 (0.41)	1.79 (0.41)	62.305 (444)	<0.01
Total	93.094 (8.75)	42.23 (7.30)	66.652 (444)	<0.01

Note. Abbreviations stand for corresponding factors in scale: CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions

Predictive Validity

The multivariate regression analysis failed to reveal that the school students' sex, age, grade, or duration of the national exam preparation were associated with GIPSS-Kz scores (**Table 7**).

Table 7. Coefficient estimates of model with four predictors & school student scores in the Kazakh version of GIPSS as an outcome variable (n=826)

Model	β (95% CI)	Beta	t-test	p
CGI subscale				
Intercept	3.487 (1.165, 5.809)	0.000	2.943	0.003
Age	-0.101 (-0.389, 0.186)	-0.075	-0.690	0.490
Grade	0.116 (-0.190, 0.421)	0.081	0.741	0.459
PP	-0.075 (-0.226, 0.077)	-0.050	-0.967	0.334
Sex	-0.026 (-0.097, 0.044)	-0.026	-0.729	0.466
OAGI subscale				
Intercept	4.203 (2.078, 6.329)	0.000	3.876	0.000
Age	-0.041 (-0.305, 0.222)	-0.034	-0.308	0.758
Grade	-0.046 (-0.326, 0.234)	-0.035	-0.322	0.747
PP	-0.006 (-0.145, 0.133)	-0.004	-0.082	0.935
Sex	0.016 (-0.049, 0.080)	0.017	0.483	0.630
AGQ subscale				
Intercept	4.789 (2.789, 6.789)	0.000	4.693	0.000
Age	-0.169 (-0.417, 0.079)	-0.146	-1.336	0.182
Grade	0.099 (-0.164, 0.363)	0.081	0.738	0.461
PP	0.116 (-0.015, 0.247)	0.090	1.741	0.082
Sex	-0.015 (-0.075, 0.046)	-0.017	-0.472	0.637

Table 7 (continued). Coefficient estimates of model with four predictors & school student scores in the Kazakh version of GIPSS as an outcome variable (n=826)

Model	β (95% CI)	Beta	t-test	p
AGI subscale				
Intercept	5.164 (3.156, 7.171)	0.000	5.042	0.000
Age	-0.231 (-0.479, 0.018)	-0.198	-1.817	0.070
Grade	0.162 (-0.102, 0.427)	0.132	1.204	0.229
PP	0.121 (-0.011, 0.252)	0.093	1.803	0.072
Sex	0.010 (-0.051, 0.071)	0.011	0.315	0.753
PGQ subscale				
Intercept	5.406 (3.259, 7.552)	0.000	4.936	0.000
Age	-0.161 (-0.427, 0.105)	-0.129	-1.189	0.235
Grade	0.015 (-0.267, 0.298)	0.112	0.107	0.915
PP	0.036 (-0.104, 0.176)	0.026	0.504	0.615
Sex	0.018 (-0.047, 0.083)	0.019	0.537	0.591
Overall				
Intercept	5.406 (3.259, 7.552)	0.000	4.936	0.000
Age	-0.161 (-0.427, 0.105)	-0.129	-1.189	0.235
Grade	0.015 (-0.267, 0.298)	0.012	0.107	0.915
PP	0.036 (-0.104, 0.176)	0.026	0.504	0.615
Sex	0.018 (-0.047, 0.083)	0.019	0.537	0.591

Note. β: Unstandardized regression coefficient; CI: Confidence interval; Beta: Standardized regression coefficient; PP: Preparation period; CGI: Communicating geographic information; OAGI: Organizing & analyzing geographic information; AGQ: Answering geographic questions; AGI: Acquiring geographic information; & PGQ: Posing geographic questions

External Consistency Reliability

The resulting intraclass correlation coefficient for the difference between first and second tool administration was 0.53 (0.40, 0.59), indicating moderate agreement.

Bland-Altman analysis yielded mean difference between scale scores at time 1 and time 2 equal to 0.22 (**Figure 3**), which is below defined top margin of 0.29. Coefficient of reproducibility obtained was 9.5%. Altogether, test-retest reliability evaluation suggests GIPSS-Kz has quite a high level of reproducibility.

DISCUSSION

This investigation is first to translate the GIPSS questionnaire into Kazakh and psychometrically validate it on a representative Kazakh-speaking population. The resultant GIPSS-Kz met the standard requirements in terms of predictive and construct validity, internal consistency, external reliability, and discrimination capability. Culturally and linguistically adapting the measurement while preserving the original structure facilitates the comparisons within and across countries (Jorgensen et al. 2020). Findings from statistical modeling in this study support replicating the prototype scale structure in the Kazakh version. For example, the percentage of variance explained by the model obtained herein enables us to assume that working with the five-

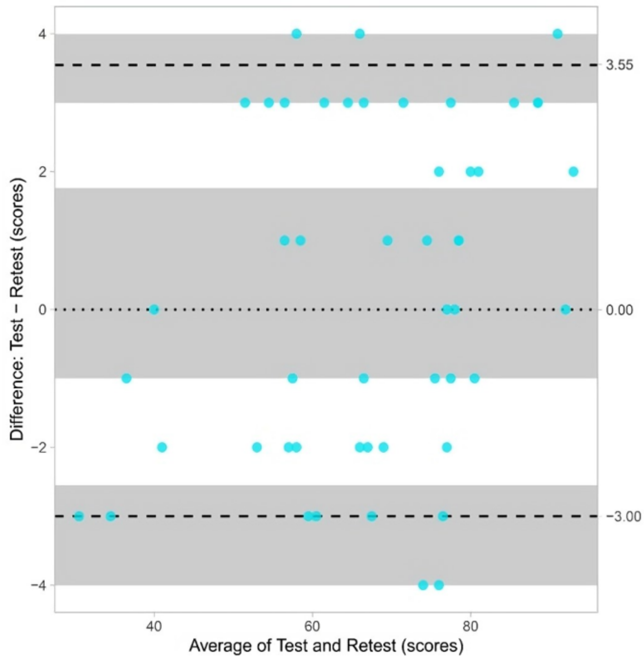

Figure 3. Bland-Altman plot for test & retest scores of the Kazakh version of GIPSS (n=50); dots represent individual data; dotted line indicates average difference between test & retest (0.22); dashed lines indicate limits of agreement (-3.0, 3.55); gray areas indicate 95% confidence intervals; & Shapiro-Wilk test of difference retrieved a p-value of 0.0023, reflecting a non-normal distribution, so limits of agreement were drawn from non-parametric estimation (Source: Authors' own elaboration)

factor model is adequate. However, the model is susceptible to improvement. Particularly, significant positive correlations between the subscales were revealed, which may argue in favor of a unidimensional structure model for the scale (Padron et al., 2012), so oncoming studies are proposed to further analyze factorial invariance.

The five-factor solution reflects the five sub-skills of scientific inquiry, which can be roughly envisaged as the steps of geographical inquiry performed by a Kazakhstani secondary school student, as follows:

1. **PGQ subscale:** Recognizing an issue, e.g., the ecological degeneration of Lake Balkhash in Kazakhstan.
2. **AGI subscale:** Collecting information on the subject of interest and excogitating the inquiry algorithm, e.g., gathering data on water levels in Lake Balkhash, etc. This stage somewhat overlaps the next one.
3. **OAGI subscale:** Reading, processing and interpreting the geographic information, e.g., juxtaposing satellite images of Lake Balkhash taken in the 1990s and nowadays.
4. **AGQ subscale:** Making conclusions and predictions based on the results of the evaluation, e.g., "the shrinking of Lake Balkhash is greatly related to a 2.5-fold water level decline in the Ili River compared to 1993; if the problem is ignored, the river may dry completely in another 30 years, which would be a catastrophe to the Balkhash ecosystem."
5. **CGI subscale:** Conveying the findings to an audience appropriately (let us take a PowerPoint presentation as the simplest example).

As such, the tool looks meaningful, practicable, and straightforward to interpret. The present survey thus provided an effective modified measure for future related studies in a specific Kazakhstani context.

This work adds to the initial study in the way that potential moderators were tested here, i.e., gender, grade, age, and time spent preparing for the geography exam. Ultimately, nonsignificant associations were found between these variables and GIPSS-Kz scores. This runs counter to findings in a research on 780 secondary school students (Nehring et al., 2015), which found that gender and grade significantly predicted scientific inquiry skills measured with a 27-item test. Plus, a study involving 145 grade 11 students (Wang et al., 2015) concluded there were statistically significant gender differences in post-test scores for some components of inquiry skills. Yet the evidence from an experiment on 88 secondary learners (Pedaste & Altin, 2020) suggests that the pupils' gender did not influence improvement in their inquiry learning as per an 11-item test. Future direction for research on the topic might be to examine potential covariates including those mentioned in Nehring et al. (2015), that is migration background, social status, spoken languages at home, and motivational variables such as a specific interest in the content.

As regards limitations of this research, our findings are based on a sample obtained from high school students in Kazakhstan within a specific age range. Therefore, these findings may not generalize to other populations or more diverse groups. One more limitation of this study resides in its self-reporting nature of the data. We recommend that future research should estimate convergent validity of GIPSS-Kz with some test employing objective criteria to measure geoinquiry skills. In this regard, the inquiry skills assessment for secondary-level-schoolers devised by Jeskova et al. (2018) seems to us to be the closest to the ideal among those publicly available. But it covers computer science, physics, and mathematics alongside geography, so comparability would be a hassle. Certainly, it would be enlightening to juxtapose the findings of the survey with the results of the cross-country geography assessment. But first, considerable portion of the respondents had not yet passed the exam by the time this paper was written; second, whichever we choose as the source of data on the scores obtained by the students would entail hindrances since the Ministry of Education would refuse to share the data with third parties, whereas the results communicated directly by the schoolers could still not be deemed unbiased.

It is hard to provide an extensive discussion comparing our findings with the past evidence owing to a lack of similar research. On the other hand, this underscores the novelty of this survey. The form has not been utilized in Kazakhstan heretofore, and we expect its validation will contribute towards that. Hopefully, the GIPSS-Kz will grease the wheels for more informed pedagogical decisions in schools.

CONCLUSIONS

Taken together, this validation study maintains the five-dimensional structure with the 22 items that make up the GIPSS. The Kazakh version possesses acceptable factorial validity and good reliability indicators. Findings from the test-retest reliability evaluation indicate the scale as consistent over a two-week measurement interval, affirming its accuracy and stability. The newly emerged questionnaire is relatively brief and simple to administer and score. In this sense, it seems feasible for assessing geographic inquiry skills in secondary school level. Therefore, the Kazakh version of GIPSS could be a valuable scale applicable in other ethnically Kazakh settings, in particular to map, where secondary students' inquiry skills are the weakest, which would allow for prioritizing, where to intervene. Moreover, adapting GIPSS to new linguacultural contexts is warranted to appraise its potential usefulness worldwide.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Acknowledgements: The authors would like to thank Sergei Kozhevnikov for helpful English correction.

Ethical statement: GIPSS designers (Ozudogru & Demiralp, 2022) were contacted by the corresponding author via e-mail in mid-November 2022, and they granted permission to employ the questionnaire. The Research Ethics Board at the corresponding author's institution approved this study to be undertaken (approval number IRB-544, 21 November 2022).

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Al Mamun, M. A., Lawrie, G., & Wright, T. (2022). Exploration of learner-content interactions and learning approaches: The role of guided inquiry in the self-directed online environments. *Computers & Education, 178*, 104398. https://doi.org/10.1016/j.compedu.2021.104398

Aldahmash, A. H., & Omar, S. H. (2021). Analysis of activities included in Saudi Arabian chemistry textbooks for the inclusion of argumentation-driven inquiry skills. *Studies in Educational Evaluation, 68*, 68100968. https://doi.org/10.1016/j.stueduc.2020.100968

Avelino, P. R., Nascimento, L. R., Menezes, K. K., Alvarenga, M. T. M., Fortini, I. F., & Teixeira-Salmela, L. F. (2022). Test-retest reliability and measurement error of the modified gait efficacy scale in individuals with stroke. *Physiotherapy Theory and Practice, 38*(13), 1-6. https://doi.org/10.1080/09593985.2021.1952669

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.

Baum, E. J. (2013). Augmenting guided-inquiry learning with a blended classroom approach. *Journal of College Science Teaching, 42*(6), 27-33. https://doi.org/10.2505/4/jcst13_042_06_27

Bhat, N. A., Devdutt, J., Johnson, J. A., & Roopesh, B. N. (2021). Adaptation and psychometric validation of Hindi version of the behavioral emotion regulation questionnaire. *Asian Journal of Psychiatry, 62*, 102730. https://doi.org/10.1016/j.ajp.2021.102730

Bidell, T. R., & Fischer, K. W. (1992). Beyond the stage debate: Action, structure, and variability in Piagetian theory and research. In R. J. Sternberg, & C. A. Berg (Eds.), *Intellectual development*. Cambridge University Press.

Brenner, P. S., & DeLamater, J. (2016). Lies, damned lies, and survey self-reports? Identity as a cause of measurement bias. *Social Psychology Quarterly, 79*(4), 333-354. https://doi.org/10.1177/0190272516628298

Curtis, M. D. (2019). Professional technologies in schools: The role of pedagogical knowledge in teaching with geospatial technologies. *Journal of Geography, 118*(3), 130-142. https://doi.org/10.1080/00221341.2018.1544267

Darmuki, A., Nugrahani, F., Fathurohman, I., Kanzunnudin, M., & Hidayati, N. A. (2023). The impact of inquiry collaboration project based learning model of Indonesian language course achievement. *International Journal of Instruction, 16*(2), 247-266. https://doi.org/10.29333/iji.2023.16215a

De La Torre, S., Spruijt-Metz, D., & Farias, A. J. (2022). Survivors' health competence mediates the association between wearable activity tracker use and self-rated health: HINTS analysis. *Journal of Cancer Survivorship, 16*(6), 1268-1278. https://doi.org/10.1007/s11764-021-01112-9

Dilbaz, N., Noyan, O., Alpar, G., & Kazan, O. (2020). Development of the COVID-19 phobia scale: Validity and reliability study. *The Journal of Neurobehavioral Sciences, 7*(3), 142-151. https://doi.org/10.4103/jnbs.jnbs_27_20

Ding, X., & Vancleef, K. (2022). Test-retest reliability and practice effect of the Leuven perceptual organization screening test. *Behavior Research Methods, 54*(5), 2457-2462. https://doi.org/10.3758/s13428-021-01741-z

Eze, E. (2021). Why secondary school geography students perform poorly in external examinations. *Journal of Geography, 120*(2), 51-60. https://doi.org/10.1080/00221341.2020.1860114

Feyzioglu, B. (2019). The role of inquiry-based self-efficacy, achievement goal orientation, and learning strategies on secondary-school students' inquiry skills. *Research in Science & Technological Education, 37*(3), 366-392. https://doi.org/10.1080/02635143.2019.1579187

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research, 82*(3), 300-329. https://doi.org/10.3102/0034654312457206

Ganajova, M., Sotakova, I., Lukač, S., Jeskova, Z., Jurkova, V., & Orosova, R. (2021). Formative assessment as a tool to enhance the development of inquiry skills in science education. *Journal of Baltic Science Education, 20*(2), 204-222. https://doi.org/10.33225/jbse/21.20.204

Gao, Y., Dai, H., Jia, G., Liang, C., Tong, T., Zhang, Z., Song, R., Wang, Q., & Zhu, Y. (2020). Translation of the Chinese version of the nomophobia questionnaire and its validation among college students: Factor analysis. *JMIR MHealth and UHealth, 8*(3), e13561. https://doi.org/10.2196/13561

Gillon, K., & Stotter, J. (2011). Inquiry learning with senior secondary students: Yes it can be done! *Access, 25*(3), 14-19. https://doi.org/10.29173/iasl7744

Gonzalez, R. D. M., & Torres, M. L. D. L. (2020). WebGIS implementation and effectiveness in secondary education using the digital atlas for schools. *Journal of Geography, 119*(2), 74-85. https://doi.org/10.1080/00221341.2020.1726991

Harte, W., & Reitano, P. (2016). "Doing geography": Evaluating an independent geographic inquiry assessment task in an initial teacher education program. *Journal of Geography, 115*(6), 233-243. https://doi.org/10.1080/00221341.2016.1175496

Heindl, M. (2019). Inquiry-based learning and the pre-requisite for its use in science at school: A meta-analysis. *Journal of Pedagogical Research, 3*(2), 52-61. https://doi.org/10.33902/jpr.2019254160

Hohnle, S., Fögele, J., Mehren, R., & Schubert, J. C. (2015). GIS teacher training: Empirically-based indicators of effectiveness. *Journal of Geography, 115*(1), 12-23. https://doi.org/10.1080/00221341.2015.1016546

Hussien, S., Wahab, A. M. K., & Hashim, R. (2021). Improving students' inquiry skills in Islamic education through hikmah pedagogy and community of inquiry. *Malaysian Journal of Learning and Instruction, 18*(2), 189-214. https://doi.org/10.32890/mjli2021.18.2.7

Jant, E. A., Uttal, D. H., Kolvoord, R., James, K., & Msall, C. (2019). Defining and measuring the influences of GIS-based instruction on students' STEM-relevant reasoning. *Journal of Geography, 119*(1), 22-31. https://doi.org/10.1080/00221341.2019.1676819

Jesková, Z., Balogova, B., & Kires, M. (2018). Assessing inquiry skills of upper secondary school students. *Journal of Physics: Conference Series, 1076*, 012022. https://doi.org/10.1088/1742-6596/1076/1/012022

Jeskova, Z., Lukač, S., Snajder, U., Gunis, J., Klein, D., & Kires, M. (2022). Active learning in STEM education with regard to the development of inquiry skills. *Education Sciences, 12*(10), 686. https://doi.org/10.3390/educsci12100686

Jorgensen, J. E., Fearon, A. M., Molgaard, C. M., Kristinsson, J., & Andreasen, J. (2020). Translation, validation and test-retest reliability of the VISA-G patient-reported outcome tool into Danish (VISA-G.DK). *PeerJ, 8*, e8724. https://doi.org/10.7717/peerj.8724

Kachergis, G., Rhodes, M., & Gureckis, T. (2017). Desirable difficulties during the development of active inquiry skills. *Cognition, 166*, 407-417. https://doi.org/10.1016/j.cognition.2017.05.021

Kanbay, Y., Firat, M., Akcam, A., Cinar, S., & Ozbay, Z. (2021). Development of Firat netlessphobia scale and investigation of its psychometric properties. *Perspectives in Psychiatric Care, 58*(4), 1258-1266. https://doi.org/10.1111/ppc.12924

Kapici, H.O., Akcay, H., & de Jong, T. (2019). Using hands-on and virtual laboratories alone or together–which works better for acquiring knowledge and skills? *Journal of Science Education and Technology, 28*, 231-250. https://doi.org/10.1007/s10956-018-9762-0

Keefer, A., Singh, V., Kalb, L. G., Mazefsky, C. A., & Vasa, R. A. (2019). Investigating the factor structure of the child behavior checklist dysregulation profile in children and adolescents with autism spectrum disorder. *Autism Research, 13*(3), 436-443. https://doi.org/10.1002/aur.2233

Khalaf, B. K., & Zin, Z. B. M. (2018). Traditional and inquiry-based learning pedagogy: A systematic critical review. *International Journal of Instruction, 11*(4), 545-564. https://doi.org/10.12973/iji.2018.11434a

Kim, M., & Bednarz, R. (2013). Effects of a GIS course on self-assessment of spatial habits of mind (SHOM). *Journal of Geography, 112*(4), 165-177. https://doi.org/10.1080/00221341.2012.684356

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1-48. https://doi.org/10.1207/s15516709cog1201_1

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kremer, K., Specht, C., Urhahne, D., & Mayer, J. (2013). The relationship in biology between the nature of science and scientific inquiry. *Journal of Biological Education, 48*(1), 1-8. https://doi.org/10.1080/00219266.2013.788541

Kuang, X., Eysink, T. H. S., & de Jong, T. (2022). Effects of providing domain information on facilitating hypothesis generation in inquiry learning. *The Journal of Educational Research, 115*(5), 285-297. https://doi.org/10.1080/00220671.2022.2124219

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction, 26*(4), 512-559. https://doi.org/10.1080/07370000802391745

Lam, E. T. C., Zhang, J. J., & Jensen, B. E. (2005). Service quality assessment scale (SQAS): An instrument for evaluating service quality of health-fitness clubs. *Measurement in Physical Education and Exercise Science, 9*(2), 79-111. https://doi.org/10.1207/s15327841mpee0902_2

Lane, R., & Bourke, T. (2017). Assessment in geography education: A systematic review. *International Research in Geographical and Environmental Education, 28*(1), 22-36. https://doi.org/10.1080/10382046.2017.1385348

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning. *Review of Educational Research, 86*(3), 681-718. https://doi.org/10.3102/0034654315627366

Lehtinen, A., Nieminen, P., Pehkonen, S., & Hahkiöniemi, M. (2022). Comparing guidance via implicit and explicit model progressions in a collaborative inquiry-based learning environment with different-aged learners. *Education Sciences, 12*(6), 393. https://doi.org/10.3390/educsci12060393

Machado, P. P., Grilo, C. M., Rodrigues, T. F., Vaz, A. R., & Crosby, R. D. (2020). Eating disorder examination–questionnaire short forms: A comparison. *International Journal of Eating Disorders, 53*(6), 937-944. https://doi.org/10.1002/eat.23275

Madan, D. B., & Wang, K. (2021). Option implied vix, skew and kurtosis term structures. *International Journal of Theoretical and Applied Finance, 24*(5), 2150030. https://doi.org/10.1142/s0219024921500308

Maddox, L. E., Howell, J. B., & Saye, J. W. (2018). Designing geographic inquiry: Preparing secondary students for citizenship. *Journal of Geography, 117*(6), 254-268. https://doi.org/10.1080/00221341.2018.1495249

Marcello, J. S. (2009). A proposal for assessment in geography education. *Journal of Geography, 108*(4-5), 226-232. https://doi.org/10.1080/00221340902946006

Metoyer, S., & Bednarz, R. (2016). Spatial thinking assists geographic thinking: Evidence from a study exploring the effects of geospatial technology. *Journal of Geography, 116*(1), 20-33. https://doi.org/10.1080/00221341.2016.1175495

Moreno, A., & Bartolome, E. (2021). Study and research paths to improve web-based inquiry learning: Study case of an ICT course in engineering. *Education Sciences, 11*(12), 772. https://doi.org/10.3390/educsci11120772

Nehring, A., Nowak, K. H., Belzen, A. U., & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education, 37*(9), 1343-1363. https://doi.org/10.1080/09500693.2015.1035358

Oberle, A. (2020). Advancing students' abilities through the geo-inquiry process. *Journal of Geography, 119*(2), 43-54. https://doi.org/10.1080/00221341.2019.1698641

Oda, K., Herman, T., & Hasan, A. (2019). Properties and impacts of TPACK-based GIS professional development for in-service teachers. *International Research in Geographical and Environmental Education, 29*(1), 40-54. https://doi.org/10.1080/10382046.2019.1657675

Osborne, Z. M., van de Gevel, S. L., Eck, M. A., & Sugg, M. (2019). An assessment of geospatial technology integration in K–12 education. *Journal of Geography, 119*(1), 12-21. https://doi.org/10.1080/00221341.2019.1640271

Ozturk, B., Kaya, M., & Demir, M. (2022). Does inquiry-based learning model improve learning outcomes? A second-order meta-analysis. *Journal of Pedagogical Research, 6*(4), 201-216. https://doi.org/10.33902/jpr.202217481

Ozudogru, Y. H., & Demiralp, N. (2022). Developing a geographic inquiry process skills scale. *Education Inquiry, 13*(3), 374-394. https://doi.org/10.1080/20004508.2020.1864883

Padron, A., Galán, I., Durban, M., Gandarillas, A., & Rodriguez-Artalejo, F. (2012). Confirmatory factor analysis of the general health questionnaire (GHQ-12) in Spanish adolescents. *Quality of Life Research, 21*(7), 1291-1298. https://doi.org/10.1007/s11136-011-0038-x

Panaretos, D., Tzavelas, G., Vamvakari, M., & Panagiotakos, D. (2019). Investigating the role of orthogonal and non–orthogonal rotation in multivariate factor analysis, in regard to the repeatability of the extracted factors: A simulation study. *Communications in Statistics–Simulation and Computation, 48*(7), 2165-2176. https://doi.org/10.1080/03610918.2018.1435803

Pedaste, M., & Altin, H. (2020). Does inquiry-based education using robots have an effect on learners' inquiry skills, subject knowledge and skills, and motivation? *International Journal on Advanced Science, Engineering and Information Technology, 10*(4), 1403. https://doi.org/10.18517/ijaseit.10.4.12766

Pedaste, M., Maeots, M., Leijen, A., & Sarapuu, T. (2012). Improving students' inquiry skills through reflection and self-regulation scaffolds. *Technology, Instruction, Cognition and Learning, 9*(1-2), 81-95.

Perugini, S., & Bodzin, A. M. (2020). Using web-based GIS to assess students' geospatial knowledge of hurricanes and spatial habits of mind. *Journal of Geography, 119*(2), 63-73. https://doi.org/10.1080/00221341.2019.1710764

Petermann, V., & Vorholzer, A. (2022). Relationship between beliefs of teachers about and their use of explicit instruction when fostering students' scientific inquiry competencies. *Education Sciences, 12*(9), 593. https://doi.org/10.3390/educsci12090593

Petridou, E., Molohidis, A., & Hatzikraniotis, E. (2022). Assessing students' ability to apply the control of variables strategy when engaged with inquiry-based worksheets during the COVID-19 era. *Education Sciences, 12*(10), 668. https://doi.org/10.3390/educsci12100668

Preskill, H., & Torres, R. T. (1999). Building capacity for organizational learning through evaluative inquiry. *Evaluation, 5*(1), 42-60. https://doi.org/10.1177/135638909900500104

Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and validation of the student-educator negotiated critical thinking dispositions scale (SENCTDS). *Thinking Skills and Creativity, 38*, 100710. https://doi.org/10.1016/j.tsc.2020.100710

Rafiq, A. A., Triyono, M. B., & Djatmiko, I. W. (2023). The integration of inquiry and problem-based learning and its impact on increasing the vocational student involvement. *International Journal of Instruction, 16*(1), 659-684. https://doi.org/10.29333/iji.2023.16137a

Sakaluk, J. K., & Short, S. D. (2016). A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. *The Journal of Sex Research, 54*(1), 1-9. https://doi.org/10.1080/00224499.2015.1137538

Sarstedt, M., & Mooi, E. (2014). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*. Springer. https://doi.org/10.1007/978-3-642-53965-7_8

Schlemper, M. B., Athreya, B., Czajkowski, K., Stewart, V. C., & Shetty, S. (2018). Teaching spatial thinking and geospatial technologies through citizen mapping and problem-based inquiry in grades 7-12. *Journal of Geography, 118*(1), 21-34. https://doi.org/10.1080/00221341.2018.1501083

Schultz, R. B., & DeMers, M. N. (2020). Transitioning from emergency remote learning to deep online learning experiences in geography education. *Journal of Geography, 119*(5), 142-146. https://doi.org/10.1080/00221341.2020.1813791

Senisum, M., Susilo, H., Suwono, H., & Ibrohim, N. F. (2022). GIReSiMCo: A learning model to scaffold students' science process skills and biology cognitive learning outcomes. *Education Sciences, 12*(4), 228. https://doi.org/10.3390/educsci12040228

Seow, T., Chang, J., & Irvine, K. N. (2019). Field-based inquiry as a signature pedagogy for geography in Singapore. *Journal of Geography, 118*(6), 227-237. https://doi.org/10.1080/00221341.2018.1561740

Soares, J., Taveira, M. D. C., Barroso, P., & Silva, A. D. (2022). Career adapt-abilities scale-short form: Validation among Portuguese university students and workers. *Journal of Career Assessment*. https://doi.org/10.1177/10690727221129281

Solem, M., Vaughan, P., Savage, C., & De Nadai, A. S. (2021). Student- and school-level predictors of geography achievement in the United States, 1994-2018. *Journal of Geography, 120*(6), 201-211. https://doi.org/10.1080/00221341.2021.2000009

Spires, H. A., Himes, M. P., & Krupa, E. (2022). Supporting students' science content knowledge and motivation through project-based inquiry (PBI) global in a cross-school collaboration. *Education Sciences, 12*(6), 412. https://doi.org/10.3390/educsci12060412

Tak, C. C., Zulnaidi, H., & Eu, L. K. (2022). Measurement model testing: Adaption of self-efficacy and metacognitive awareness among university students. *EURASIA Journal of Mathematics, Science and Technology Education, 18*(9), em2153. https://doi.org/10.29333/ejmste/12366

Timms, M. J., Moyle, K., Weldon, P. R., & Mitchell, P. (2018). Challenges in STEM learning in Australian schools: Literature and policy review. *Australian Council for Educational Research*. https://research.acer.edu.au/policy_analysis_misc/28

Towne, F. S. (2009). *Is adolescence a critical period for learning formal thinking skills? A case study investigating the development of formal thinking skills in a short-term inquiry-based intervention program* [PhD dissertation, University of Montana].

Turner, R. C., & Carlson, L. (2003). Indexes of item-objective congruence for multidimensional items. *International Journal of Testing, 3*(2), 163-171. https://doi.org/10.1207/S15327574IJT0302_5

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS ONE, 8*(9), e73990. https://doi.org/10.1371/journal.pone.0073990

Wan Hassan, W. N., Yusof, Z. Y. M., Shahidan, S. S. Z., Mohd Ali, S. F., & Makhbul, M. Z. M. (2017). Validation and reliability of the translated Malay version of the psychosocial impact of dental aesthetics questionnaire for adolescents. *Health and Quality of Life Outcomes, 15*, 1. https://doi.org/10.1186/s12955-017-0600-5

Wang, J., Guo, D., & Jou, M. (2015). A study on the effects of model-based inquiry pedagogy on students' inquiry skills in a virtual physics lab. *Computers in Human Behavior, 49*, 658-669. https://doi.org/10.1016/j.chb.2015.01.043

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2009). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching, 47*(3), 276-301. https://doi.org/10.1002/tea.20329

Wu, H., & Hsieh, C. (2006). Developing sixth graders' inquiry skills to construct explanations in inquiry-based learning environments. *International Journal of Science Education, 28*(11), 1289-1313. https://doi.org/10.1080/09500690600621035

Xuan, X., Jin, Q., Jo, I., Duan, Y., & Kim, M. (2019). The potential contribution of geography curriculum to scientific literacy. *Journal of Geography, 118*(5), 185-196. https://doi.org/10.1080/00221341.2019.1611906

Zhdanov, S. P., Ishmuradova, A. M., Zakharova, V. L., Belous, S. V., & Grishnova, E. E. (2022). Validation and adaptation of the questionnaire on science motivation in the Russian context. *EURASIA Journal of Mathematics, Science and Technology Education, 18*(8), em2136. https://doi.org/10.29333/ejmste/12219

Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., & Sijtsma, K. (2019). Item-score reliability as a selection tool in test construction. *Frontiers in Psychology, 9*, 2298. https://doi.org/10.3389/fpsyg.2018.02298

https://www.ejmste.com