An Assessment Framework and Methodology for a Trends in International Geography Assessment Study (TIGAS)

Michael Solem
National Center for Research in Geography Education, USA
Joseph Stoltman
Western Michigan University, USA
Rod Lane
Macquarie University, Australia
Terri Bourke
Queensland University of Technology, Australia
Chew Hung Chang
National Institute of Education, Singapore
Kathrin Viehrig
University of Applied Sciences and Arts Northwestern Switzerland

Abstract

Since 2016 an international research process has been underway to design and develop an international geography assessment for implementation in lower secondary education settings. One of the crucial steps in this process is the development and validation of an assessment framework that models the content and cognitive dimensions of geography education to enable internationally valid, reliable, and fair measures of geographic constructs. This paper provides a rationale for an international assessment in geography and reports the findings of foundational research that produced the provisional assessment framework. Our methodology draws on the evidence-centered design to educational assessment development, which involves a sequential approach to domain analysis and modelling. The framework will guide the specifications for tasks and tests, evaluation procedures, and measurement models. The article concludes with a reiteration of the value of an international assessment and an outline of the activities moving forward.

Introduction

The authors are members of a study group established in 2016 to design and develop a Trends in International Geography Assessment Study (TIGAS). The idea for TIGAS originated in April 2014, when Hans Wagemaker, an evaluation consultant with International Association for the Evaluation of Educational Achievement (IEA), visited Professor Joseph Stoltman at Western Michigan University. IEA coordinates the international administration of Trends in International Mathematics and Science Study (TIMSS) and other international comparative assessments including the Progress in International Reading Literacy Study (PIRLS), International Civic and Citizenship Education Study (ICCS), and the International Computer and Information Literacy Study (ICILS).

Conversations between Wagemaker and Stoltman led to a recognition that geography would be a prime candidate for an IEA assessment. That meeting was followed with a proposal to the IGU Commission on Geographical Education (IGU-CGE) in Krakow, which approved the formation of a Task Force (Dr Rod Lane, Dr Terri Bourke, and Professor Joseph Stoltman) charged with studying the feasibility of an international geography assessment. Lane and Bourke were assigned to complete a needs/interest survey for a geography assessment modelled on TIMSS; this survey confirmed the strong interest of the international academic geography education community (Lane and Bourke, 2016a). The findings of the survey were delivered to IEA and discussed with the TIMSS, IGU-CGE, U.S., Asian, and European constituents to ascertain the best grade/age level for an international assessment. Based on information gathered from this process, IEA concluded that an assessment for lower secondary education (learners aged 13–14 years) would be the most viable option for an international geography assessment.

Rationale and Significance

There are many reasons why the proposed international geography assessment is needed. First, the assessment should encourage thinking about geography education in an international context. International collaboration is a major driver of research and discovery in the geography discipline, yet educational research in the field is almost always conducted in a national rather than an international context (Lane & Bourke, 2017). Current national, state, and provincial geography assessments were largely developed to inform domestic educational priorities, and the data they produce tend to focus on localised or regional content knowledge lacking global relevancy. TIGAS will produce an assessment that is consistent with the transnational epistemic qualities of the discipline. The project's research, design, and development activities will identify and measure geographic content and practices that represent academic outcomes that all young people need for understanding issues and processes operating across multiple scales of society and the environment (e.g., global climate change, natural hazards such as earthquakes and hurricanes, the impact of urbanisation on the availability of resources, the effects of industrial pollution on ocean ecosystems, human migration, and globalisation).

An international geography assessment would be expected to capture the depth of students' ability to think geographically beyond their local or national perspectives. As with other sciences, the nature of geographic knowledge is conceptual, theoretical, and contested. An international assessment of students' use of geographic information, facts, concepts, processes, and models is necessary to reveal how geography is understood and practised by students within diverse global contexts. This is important because no single country can resolve issues, such as global climate change, facing the world's people, places, and environments. It is unquestionably in the national interest of all countries for students to learn geography at a world standard and for educational policymakers to work cooperatively with other countries to raise the quality of geography education internationally. From this undertaking, participating countries stand to gain a citizenry capable of making more informed decisions and a workforce possessing the geographic knowledge, skills, and abilities needed to address global issues through multilateral action.

Over time, the trend data from the international geography assessment may well facilitate the development of new theories of geography learning by supporting investigations that are at present difficult or impossible to conduct, including questions such as:

- How is geography education implemented in participating countries?
- What is the extent and variation of students' geography education knowledge within and across participating countries?
- What is the extent of engagement with geography education in different spheres of society and what are the related factors within and across countries?
- What beliefs do students in participating countries hold regarding geographic issues in modern society and what are the factors influencing variation in students' dispositions?
- How are schools in the participating countries organised with regard to geography education and what is the curricular association with students' learning outcomes?
- Which beliefs do teachers in the participating countries hold regarding geographic education?

Having comparative data from an international geography assessment is the best possible way of evaluating the future capacity of students to engage internationally with the perspectives of their peers and participate as globally-minded individuals able to work cooperatively and collaboratively on issues that threaten Earth's diverse environments. Yet unlike subjects such as mathematics and science, there is currently no reliable international source of assessment data informing policymakers about what students in lower secondary school know and are able to do in geography that will help them live productive and informed lives when they complete school.

It is true that some widely recognised geographic concepts and knowledge are currently present in TIMSS Earth Science and Biology topics (e.g., weather patterns, natural resources, and anthropogenic changes to natural environments). However, many of these earth science and biology items lack a spatial or geographical context. An international assessment in geography would capture those elements of human geography that are at present not a focus of existing international assessments.

In many countries, geography lessons receive less attention in favour of subjects that are tested in international comparative studies such as the Programme for International Student Assessment (PISA) and TIMSS. Although these international assessments have been critically debated (Lane & Bourke, 2016b), it cannot be denied that they generate a strong scientific and societal source of information that is valuable for planning, policy formulation, and researching the relationship between school curricula and society.

In practice, we acknowledge that some observers may simply use the comparative data from an international assessment to rank nations on the basis of student achievement. Even so, the intent and deeper value of this project will come in the form of long-term trend studies informing what all nations must do to elevate the capabilities of students to take on the shared challenges of this day and age.

Evidence-Centered Design for TIGAS Development

TIGAS will require an assessment framework developed as a data-based document, relying on both quantitative and qualitative data from literature reviews and surveys. An evidence-centered design (ECD) for substantiating the framework theoretically is required. ECD is a structured approach to assessment development that views assessment as an evidentiary argument of what students know and can do (Brennan, 2010). The overarching research question is: *What characteristics of assessment design, implementation, and delivery enable internationally valid comparisons of what students in lower secondary education settings know and are able to do in geography?* This question is fundamental to the further development of TIGAS. Each step will be informed by ECD, which has five components with distinct roles in a comprehensive assessment process (Mislevy & Haertel, 2006).

a. ***Domain Analysis:*** This involves gathering substantive information about the domain to be assessed. For TIGAS, domain analysis will involve an international comparative analysis of geography curricula in lower secondary education with a broader sample to identify threads of geographic content and practices which extend the preliminary survey.

b. ***Domain Modelling:*** This step involves expressing the assessment argument in narrative form based on the domain analysis. We anticipate this narrative will specify geographic content and practices that the participating countries expect of students in lower secondary education.

c. ***Assessment Framework:*** Following domain analysis and modelling, the provisional framework will be further developed, expressing the assessment argument in structures and specifications for items and tasks, evaluation procedures, and measurement models. This framework will include geographic content and cognitive dimensions.

d. ***Assessment Implementation:*** This step is designed to implement the assessment including presentation-ready items and tasks and calibrated measurement models. Implementation of the international geography assessment will be conducted digitally to support items which utilise geo-visualisation and mapping technologies.

e. ***Assessment Delivery:*** The final activity in ECD involves coordinating the interaction of students with items and tasks, followed by assessment scoring, and reporting. It is planned for the IEA to manage this coordination within and between participating countries.

As the ECD model is followed, national assessments from different countries will provide sources from which to assemble prototype items for field trials and subsequent analysis. *The International Charter on Geographical Education* (IGU-CGE, 2016) served as a major source for the identification of assessable content for the assessment framework. The Charter was developed by the international community of geography educators, and thus is a definitive statement regarding geography education content internationally.

Assessment prototypes will need to meet criteria established in the assessment framework and serve as models for constructing equivalent items. Prototype items will be sourced from a variety of geography-related assessments and publications produced in different countries. These include TIMSS/PISA Geoscience subtests; The International Assessment of Educational Progress (U.S. Department of Education et al., 1992); the IGU InterGeo Project (Lambert & Purnell, 1994; Niemz & Stoltman, 1993); Global Geographic Literacy Study (National Geographic Education Foundation, 2002) and research on geographic educational assessment by Gerber (2001). Furthermore, the research will draw from assessment literatures on spatial thinking and reasoning (e.g., Chung, Cannady, & Kremer,, 2015; Huynh & Sharpe, 2013; Lee & Bednarz, 2012) and systemic thinking (e.g., Mehren et al., 2016; Viehrig, 2015; Viehrig et al., 2017).

The geoscience and geography education communities have also produced a variety of studies dealing with learner conceptualisations, for example, the geoscience concept inventory by Libarkin and Anderson (2005) and studies regarding topics, such as water (e.g., Ben-Zvi Assaraf & Orion, 2005; Reinfried, Tempelmann, & Aeschbacher, 2012), avalanches (e.g., Rempfler, 2010), tsunamis (Etterich, 2013), cyclones (e.g., Lane & Catling, 2016; Lane & Coutts, 2012), and the polar regions (Adamina, 2008; Conrad, 2012).

Viehrig and Lane are currently exploring options to use wiki technology (http://geoconcepts.geographyteachereducator.com) to collect central results and implications of different studies and make them accessible not just for the item designers within TIGAS but also for teachers and preservice teachers.

The following is a summary of work completed by the TIGAS Study Group between September 2016 and June 2018, in preparation for Phase 1 (Assessment Framework Development). This prior work focused primarily on the first two components of Evidence-Centered Design (domain analysis and domain modelling) and involved an international curriculum survey and preliminary analyses of assessment prototypes. This work resulted in a provisional assessment framework and will inform further development of the framework and item development scheduled in Phase 1.

Summary of Work Completed (September 2016 to present)

Curriculum survey

In September 2016, a survey was conducted of the eight members of the TIGAS Group to identify geographic concepts and content threads common to 8th-grade geography classrooms in South Korea, The Netherlands, Czechia, Switzerland, Singapore, Australia, and the United States. The survey questions were:

1. What type of curriculum document is used in your country for 13/14 year olds?
2. Is geography taught as a stand-alone subject for 13/14 year olds?
3. What geographical contemporary issues are present in your curriculum document for 13/14 year olds?
4. What domains are addressed in the curriculum document for 13/14 year olds?
5. What conceptual knowledge and understanding should 13/14 year old students in your country/state have learned?
6. What skills should 13/14 year old students in your country/state have learned?
7. Which representations do students in your country/state work with by age 13/14?
8. What elements of enquiry should students in your country/state work with by age 13/14?

The following is a summary of the data analysis.

Seventy-five percent of respondents had a national curriculum document where geography was taught as a stand-alone subject in lower secondary schools. The other twenty-five percent were specific to provinces, cantons, counties, departments, or regions.

The contemporary geographical issues common within the curriculum documents in the seven countries were: urbanisation; energy supplies and management; environmental quality; hazards and disasters; global change; population dynamics/migration; sustainable development and climate change. With regard to domains, the most often cited were climatic geography; population geography; economic geography; geomorphology; urban geography and cultural geography. All were common to the seven countries. The key concepts identified in the curriculum documents are shown in Table 1 together with examples of the language indicative of the concept.

With regard to skills, the curriculum documents from the seven countries focused on making decisions, working cooperatively, solving problems, making judgements, developing generalisations, identifying questions and issues, processing, interpreting and evaluating data, and collecting and structuring information. Students were expected to work with a range of visual representations including: graphs, tables, diagrams, maps, renderings from Geographic Information Systems (GIS), and photographs. Listening skills for verbal information narratives in printed materials were deemed very important skills. Most countries required students to work with quantitative, in addition to various forms of qualitative, data including cartoons, photographs, comics, transcripts and satellite images. Finally, the elements of geographical enquiry were central to curriculum documents in each country.

Review of Select Assessment Items

Next, the TIGAS Study Group analysed selected geography assessments from the U.S., Australia, and Singapore in an effort to categorise them according to targeted ability, item characteristics, and confounding factors (Edelson, Shavelson, & Wertheim, 2013).

Targeted ability refers to the substance of what an item assesses, spanning content, skills, and cognitive ability. The focus here was on students' geographic conceptions and content applications. This involved the creation of a comparison matrix for organising the international pool of geography assessment items. The matrix described the item types, stimuli, topic areas, depth of knowledge areas, and skills represented in the assessment items. *Item characteristics* describe how an assessment task is presented to the learner, including the setting, instructions, structure, and graphical representation in the stem or answer choices. *Confounding factors* were also identified.

Table 1: Key concepts from the curriculum analysis

Concepts	Examples
Space	'Spatial phenomena', 'extend their competencies regarding spatial topics', 'orient themselves spatially', 'preserve spatial foundations', 'spatial thinking'
Region	'Recognise diversity in different regions' 'Dynamic regional diversity'
Scale	'Interested in geographical issues of local, national and global contexts' 'Deep geographical knowledge of their own locality'
Interconnection	'Deal with connections and relationships' 'Relationship of humans to their natural and shaped environment' 'Human-environment interactions' 'Combined natural science/social science' 'Physical geographical aspects combined with human geographical aspects'
Change	'Development of humans and societies – reconstruct the past from the present to get orientation for the future'
Time	'Change in human/physical processes over time'
Sustainability	'To give thought to the future' 'Sustainable development' 'Contribute to the development of an environmentally and economically sustainable and socially just world'

This included the flaws in item design that needed to be corrected because they undermined item reliability, validity or fairness. Whilst the process only involved three countries, it enabled us to develop an approach for how this can be done on a much larger scale in the future.

The TIGAS Assessment Framework

During a Swiss National Science Foundation funded workshop in Windisch, Switzerland (September 10–12, 2017) a draft of the Assessment Framework was developed. In general, the framework was adapted from the principles used in TIMSS 2011 (Mullis, Martin, Foy, & Arora, 2012) and 2015 (Jones, Wheeler, & Centurino, 2013). Consideration was given to current international research and initiatives in geographic education, such as the International Charter on Geographical Education on which the curriculum survey outlined above was based (IGU-CGE, 2016).

The geography assessment framework is organised around two domains: a content domain and a cognitive domain.

The *content domain* includes four subdomains.

1. Earth's structure, physical environments and natural systems: including weather and climate, landforms, earthquakes and volcanic activity, and ecosystems.
2. Human environments and socio-economic systems: including population and settlements, economic processes, society, identities and conflicts.
3. Human-environment interactions and systems: including human activity and its relationships with processes in the atmosphere, hydrosphere, lithosphere and biosphere.
4. The world in spatial terms: including procedural knowledge, geographic methods and skills, and using such different visual representations as maps and satellite images.

The *cognitive domain* addresses students' abilities to think, demonstrate skills, and take action geographically along three cognitive processes.

1. **Knowing:** recalling, describing and providing examples, for example, knowing geographic facts, concepts, relationships and processes.
2. **Applying:** comparing, classifying, relating, interpreting, explaining or using models by applying knowledge of geographic facts,

concepts, relationships, procedures and methods in familiar contexts or in tasks that include the information needed for students to familiarise themselves with the specific spatial context.

3. **Reasoning:** analysing, synthesising, evaluating, generalising, inquiring, and extending knowledge and understanding to new geographic contexts.

The provisional framework can be found on the TIGAS project webpage at http://www.tigas2023.com/2018/tigas-assessment-framework-draft/

To elicit feedback from the international community about the draft framework, a social lab was conducted in Lisbon at the IGU-CGE Conference. The term *social lab* is used to describe the process of bringing together a diverse group of stakeholders to create new insights and to collaboratively explore, frame and co-create solutions to complex challenges. In social labs, emphasis is placed on dialogue, listening carefully, sharing ideas and prototyping solutions. The provisional framework that resulted from the Switzerland Workshop in 2017 was presented to participants using the following process as the social lab strategy.

1. Mapping the system: participants were introduced to the draft TIGAS framework.
2. Questioning existing approaches: The social lab participants collectively reflected on the domains and subdomains in the TIGAS framework by discussing and responding to the following prompts.

    a. Do you agree with the core concepts that have been outlined to represent the content domain?
    b. Do you think that anything is missing, should be changed or deleted (think about your country's curriculum document)?
    c. The content domains are elaborated as outcomes. Do you think that the cognitive level is appropriate for Grade 8 (13–14 years old)?
    d. With reference to the content domain, do you believe that the target percentages in terms of assessment time are appropriate for Grade 8?
    e. With reference to the cognitive domain, do you believe that the target percentages in terms of assessment time are appropriate for Grade 8? (Please note that we adapted the model for the TIMSS framework for descriptions of the cognitive domain).
    f. With reference to geographic practices, do you think that the framework captures the skills fundamental to the discipline of Geography? What would you note that appears to be missing?
    g. What other feedback do you have regarding the TIGAS framework?

Discussions were audio recorded and transcribed verbatim. Analysis of the social lab transcript will inform the next iteration of the framework. Further consultation will take place on the revised document. This is a work in progress. The authors encourage readers of this paper to provide feedback on the provisional framework by completing the questionnaire located on the TIGAS webpage as a continuation of the social lab: http://www.tigas2023.com/

Draft schedule for further development

To advance the process of developing the international geography assessment, the TIGAS group has planned four phases of design and development beginning in July 2018 (Figure 1). The Assessment Development Committee will consist of geography educators and geography education researchers from the participating countries, the co-Principal Investigators, and senior personnel including the lead Educational Testing Service (ETS) assessment developer. The ETS test development team will include the lead assessment developer, two assessment developers, fairness and editorial reviewers, a psychometrician, a statistical data analyst, a research scientist, as well as administrative and information technology staff who will be responsible for preparation of the assessment forms.

The TIGAS project's research, design, and development will emulate the collaborative process managed by the TIMSS & PIRLS International Study Center at Boston College. The participation of both IEA and Boston College personnel will enhance the probability that the internationally-validated geography items being developed will be ready for presentation and acceptance when the TIMSS National Research Coordinators next meet in 2021. Secure assessment items in TIMSS appear with every iteration or are modified as necessary to reflect changes in the scope of national curricula and to ensure comparability with prior TIMSS assessments. New item development is performed by TIMSS National Research Coordinators through a process that is dedicated to ensuring that the assessment materials can be translated accurately and used to measure comparable student outcomes in mathematics, science, and literacy skills. Because the proposed geography assessment will have no current international precursors, all of the items will need to be empirically-tested in international field trials.

Figure 1. Overall Timetable for TIGAS Design and Development

Phase 1. Assessment Framework Development	Phase 2. Item Development and Pilot Form Assembly	Phase 3. Pilot Assessment	Phase 4. Analysis and Reporting
2018–2019	2019–2020	2020–2021	2021–2022
Convene Assessment Development Committee (ADC) to author and review assessment framework.	Item selection. ADC review of pilot items, ETS content, fairness, and editorial reviews of pilot items.	Preparation of assessment forms for electronic administration. Preparation of administration software training for proctors.	Preliminary item analysis. Standard setting; Factor analysis; Item Response Theory; Differential item functioning and bias review. Final item analysis.
Select prototype items from national and provincial assessments.	Assessment forms assembly. ADC review of assembled forms.	Deployment of assessment software to pilot administration sites.	Preliminary exam score reporting. Committee review of scoring, assessment report development, and pilot program evaluation.
External reviews of assessment framework.	External review of assembled forms by assessment experts.	Assessment administration at pilot sites.	Pilot assessment report dissemination.
Item writing workshop; Item writing assignments.	Translation of assessment forms from English to national languages of non-English speaking countries.	Preparation of assessment data for analysis. Scoring of constructed response items.	Preparation for international geography assessment deployment during TIMSS 2023.

Conclusions

This paper provides a rationale for an international assessment in geography and has reported the findings of foundational research that produced a provisional assessment framework for TIGAS. A schedule for the further development of the framework and TIGAS assessment has been outlined. This development process affords a number of opportunities. These include building capacity for long-term innovative assessment research in geography education. Teachers and policymakers need current research data to make informed decisions about the educational needs of young people. There will also be opportunities for advanced level graduate students and early career scholars to develop original dissertations and postdoctoral research studies in such areas of psychometric research as item response theory, factor analysis, cognitive diagnostic modelling, and differential item and assessment functioning (Price, 2016; Penfield & Camilli, 2007; DiBello, Roussos, & Stout, 2007). Data from the international geography assessment will additionally support efforts to develop geography curricula focused on social and environmental issues operating across local, national, and international scales.

Acknowledgments

Members of the TIGAS Study Group include Erik Bijsterbosch (Windesheim University of Applied Sciences, the Netherlands), Miroslav Marada (Charles University, Czechia), Jon Moore (Educational Testing Service, USA), Okkyong Yoon (Cheongju National University of Education, South Korea). Dirk Hastedt (International Association for the Evaluation of Educational Achievement) and Eugenio Gonzalez (ETS) offered important consultation. The Swiss National Science Foundation, Geography Education National Implementation Project, and National Center for Research in Geography Education provided funding for the work reported in this article.

References

Adamina, M. (2008). *Vorstellungen von Schülerinnen und Schülern zu raum-, zeit- und geschichtsbezogenen Themen: Eine explorative Studie in Klassen des 1., 3., 5. und 7. Schuljahres im Kanton Bern* [Perceptions of students on space, time and history related topics: An explorative study in classes of the 1st, 3rd, 5th and 7th school year in the canton of Bern].(Doctoral thesis, Westfälische Wilhelms-Universität, Muenster, Germany). Retrieved from https://d-nb.info/989758885/34

Ben-Zvi Assaraf, O., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, *42*, 518–560.

Brennan, R. (2010). Evidence-centered assessment design and the Advanced Placement® Program: A psychometrician's perspective. *Applied Measurement in Education*, *23*, 392–401.

Chung, J., Cannady, M., & Kremer, A. (2015). About the geospatial reasoning instrument. Unpublished report.

Conrad, D. (2012). Schülervorstellungen zur eisigen Welt der Polargebiete. Ergebnisse einer explorativ angelegten Studie [Pupils' ideas about the icy world of the polar regions: Results of an exploratory study]. *Geographie und Ihre Didaktik*, *40,* 105–127.

DiBello, L. V., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics*, *volume 26*, *Psychometrics* (Chapter 31A). Amsterdam: Elsevier.

Edelson, D. C., Shavelson, R. J., & Wertheim, J. A. (Eds.). (2013). *A road map for 21st century geography education: Assessment: Recommendations and guidelines for assessment in geography education*. Washington DC: National Geographic Society.

Etterich, M. (2013). *Tsunami – Naturereignis, Naturgefahr, Naturkatastrophe? Schülervorstellungen und wissenschaftliche Vorstellungen zu einem Thema zwischen Physischer Geographie und Humangeographie* [Tsunami – natural phenomenon, natural hazard, natural disaster? Pupils' ideas and scientific ideas on a topic between physical geography and human geography] (Master's thesis, University of Münster, Germany). Retrieved from https://miami.uni-muenster.de/Record/455bd7a4-4477-4ea8-bce4-76304f3c8da1

Gerber, R. (2001). The state of geographical education in countries around the world. *International Research in Geographical and Environmental Education*, *10*, 349–362.

Huynh, N. T., & Sharpe, B. (2013). An assessment instrument to measure geospatial thinking expertise. *Journal of Geography, 112*, 3–17.

International Geographical Union. Commission on Geographical Education [IGU-CGE]. (2016). *International charter on geographical education*. Retrieved from http://www.cnfg.fr/wp-content/uploads/2017/12/Charter_2016-IGU-CGE_May_9.pdf

Jones, L. R., Wheeler, G., & Centurino, V. A. S. (2013). TIMSS 2015 Science Framework. In I. V. S. Mullis & M. O. Jones Martin, (Eds.), TIMSS 2015 assessment frameworks (Chapter 2). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

Lambert, D., & Purnell, K. (1994). International testing in geography: comparing students' achievements within and between countries. *Assessment in Education: Principles, Policy & Practice*, *1,* 167–180.

Lane, R., & Bourke, T. (2016a). Possibilities for an international assessment in geography. *International Research in Geographical and Environmental Education*, January 2016. http://dx.doi.org/10.1080/10382046.2016.1165920

Lane, R., & Bourke, T. (2016b). The inclusion of geography in TIMSS: can consensus be reached? *International Research in Geographical and Environmental Education*, 1–11. http://dx.doi.org/10.1080/10382046.2016.1178939

Lane, R., & Bourke, T. (2017). Assessment in geography education: a systematic review. *International Research in Geographical and Environmental Education*, 1–15. https://doi.org/10.1080/10382046.2017.1385348

Lane, R., & Catling, S. (2016). Preservice primary teachers' depth and accuracy of knowledge of tropical cyclones. *Journal of Geography*, *115*, 198–211.

Lane, R., & Coutts, P. (2012). Students' alternative conceptions of tropical cyclone causes and processes. *International Research in Geographical and Environmental Education*, *21*, 205–222.

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*, 15–26.

Libarkin, J., & Anderson, S. (2005). Assessment of learning in entry-level Geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education, 53*, 394–401.

Mehren, R., Rempfler, A., Ullrich-Riedhammer, E.-M., Buchholz, J., & Hartig, J. (2016). Systemkompetenz im Geographieunterricht: Ein theoretisch hergeleitetes und empirisch überprüftes Kompetenzstrukturmodell [System competence in geography lessons: A theoretically derived and empirically verified competence structure model]. *Zeitschrift für Didaktik der Naturwissenschaften*, *22*, 147–163.

Mislevy, R., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20.

Mullis, I. V. S., Martin, M. O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Geographic Education Foundation. (2002). *National Geographic-Roper 2002 global geographic literacy survey*. Washington DC: National Geographic Society.

Niemz, G., & Stoltman, J. (1993). *InterGeo II: International Geographical Achievement Test. Field trials report and test (Secondary Schools, Grade 8)*. Retrieved from https://archive.org/stream/ERIC_ED372985/ERIC_ED372985_djvu.txt

Penfield, R., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, volume 26, Psychometrics* (Chapter 5). Amsterdam: Elsevier.

Price, L. R. (2016). *Psychometric methods: Theory into practice*. New York: Guilford Press.

Reinfried, S., Tempelmann, S., & Aeschbacher, U. (2012). Addressing secondary school students' everyday ideas about freshwater springs in order to develop an instructional tool to promote conceptual reconstruction. *Hydrology and Earth System Sciences, 16*, 1365–1377.

Rempfler, A. (2010). Fachliche und systemische Alltagsvorstellungen von Schülerinnen und Schülern zum Thema Lawinen [Professional and systemic everyday ideas of pupils about avalanches]. In S. Reinfried (Ed.), *Schülervorstellungen und geographisches Lernen: Aktuelle Conceptual-Change-Forschung und Stand der theoretischen Diskussion* [Pupil conceptions and geographic learning: Current conceptual change research and state of theoretical discussion] (pp. 55–85). Berlin: Logos-Verlag.

United States Department of Education, National Center for Education Statistics, International Assessment of Educational Progress. (1992). Table 402: Geography proficiency of 13-year-olds in educational systems participating in the International Assessment of Educational Progress: 1991, Learning about the World, http://nces.ed.gov/programs/digest/d99/tables/PDF/Table402.pdf, (2011-09-14).

Viehrig, K. (2015). *Exploring the effects of GIS use on students' achievement in geography*. Heidelberg: Heidelberg University of Education.

Viehrig, K., Siegmund, A., Funke, J., Wüstenberg, S., & Greiff, S. (2017). The Heidelberg inventory of geographic system competency model. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments*. Springer, Cham: Switzerland.